

VARIATION RESOLUTIONS FOR CMOS SENSING NETWORKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yingqiu Cao

August 2018

© 2018 Yingqiu Cao
ALL RIGHTS RESERVED

VARIATION RESOLUTIONS FOR CMOS SENSING NETWORKS

Yingqiu Cao, Ph.D.

Cornell University 2018

Variation and variability have become the main concerns for reliable design methodology in CMOS sensing networks. On one hand, the variation can originate from the sensor tag itself, where the performance can be compromised by the uncontrollable variations. Process variation is an unavoidable consequence from the continuous scaling of modern CMOS technologies, which has increased the transistor count from thousands to billions and improved the microprocessor operating frequency from MHz to GHz. Process variation will reduce the sensor tag sensitivity and distort the transduction accuracy. In this dissertation, we will use the process variation effects on the RF-to-DC rectifiers with process variation as an illustration, as many passive sensors need such units to scavenge the ambient energy to accomplish the sensing functions. To counter such process variation, in this dissertation, a novel tunable- V_{th} rectifier based on floating-gate MOS diodes is proposed and implemented in logic CMOS foundry technology. The proposed tunable- V_{th} rectifier can have each constituent diode tuned to its optimal threshold at the wafer testing stage to maximize the operational output voltage at various loads and to compensate device-level variations. An optimization algorithm was implemented to automatically take the output voltage as feedback for system calibration in a short duration. The measurements of the tunable-threshold rectifier show $> 4\text{dB}$ improvement in input sensitivity compared to the rectifier built by foundry-provided zero-threshold transistors. The proposed circuits can be combined

with other techniques including high- Q impedance matching for input voltage boosting and hierarchical tandem stages for further improvement on operating conditions. With a $Q = 10$ matching network, -27 dBm sensitivity and 22% efficiency can be achieved for about 0.5 V DC output to a 500 k Ω load at 570 MHz.

On the other hand, the sensing variation can originate from the targeted biological sensing signal, which complicates both the sensor system design and the associated signal analysis. We will illustrate a spike-sorting method to reliably classify the enteric neural signals which have unique waveform features but large variation in magnitude, timing and duration. The proposed fastDTW spike classification algorithm provides improvements in accuracy and computational cost in comparison with Cross-correlation based template matching and PCA + k-means clustering without time warping. When applied to mouse ENS neurons in high noise and high variability environment, fastDTW successfully recognized spikes with variability is as large as 1.2 ms in width and a few millivolt in magnitude. The captured waveform features are used for variation correlation analyses to better understand the operating principles of enteric nervous system.

Although other variation sources can also affect the sensor system design, our approaches of device compensation based on operational feedback and signal tolerance based on time warping are able to give illustrations for sensor designers to successfully countermeasure uncontrollable variation sources.

BIOGRAPHICAL SKETCH

Yingqiu Cao was born in Jiaxing, a small city to the south of Shanghai in Zhejiang Province, China. She attended Jiaxing No.1 high school where she explored her interests in chemistry, and Zhejiang University in Hangzhou, where she received the bachelor degree from the Department of Optical Engineering.

After several trail and attempts in different areas of science and engineering, Yingqiu finally found her true interests in Electrical and Computer Engineering, and joined the Ph.D program at Cornell in Fall 2012. During her Ph.D, her research interests lie in a wide range of fields including energy harvesting in RFID, RFIC designs, bio-electronics, and algorithm development and signal processing for biological neuron networks. In Fall 2016, Yingqiu was a research intern at IBM on the development of advanced processes in Albany, NY.

Yingqiu had six years experiences in Chinese Calligraphy and piano. She was on the track team for competitive aerobics in middle school. In her leisure time, she enjoys watching Sci-Fi movies, playing video and board games and preparing Chinese cuisine.

This document is dedicated to my parents and my Fiancé Christopher L. Torng.

ACKNOWLEDGEMENTS

My research work would have been impossible without the aid and support from my advisor, special committee members, group mates, family and friends.

First and foremost, my sincere gratitude goes to my Ph.D. advisor, Prof. Edwin C. Kan, for his motivation, patience, caring the immense knowledge. I came to the group from a background in Optical Engineering. Prof. Kan was the person who introduced me to the field of Electrical Engineering, throughout countless number of discussions and dedicated hours of tutoring. He guided me in research projects on bio-electronics and RFID systems. There were many times when my research seemed to have reached the bottleneck. But I never lost faith because I knew he was there to support and enlighten. I could not have become who I am today without his help. Besides being a great advisor, Prof. Kan is a good friend who offered a great deal of life suggestions and career advices.

I would also like to thank Prof. Alyosha C. Molnar and Prof. Amit Lal for being my special committee members. Prof. Molnar taught me analog and RF integrated circuits and provided numerous useful comments for my IC design projects. I learned my MEMS basics from Prof. Lal.

I would like to express my gratitude to the past and the current group members. I thank Krishna Jayant for being my mentor and teaching me the basic experimental skills in CMOS sensing for bio-medical applications. I thank Yunfei Ma for giving invaluable advice on my IC designs and Ph.D life. Kshitij Auluck was my first person to go to, whenever I have a problem related to device modeling and solid-state physics. Philip Gordon also helped me with my experiments on neural recording. I would also like to thank Xiaonan, who offered kind and helpful advice for my custom PCB design. I am also grateful to my other group members: Sarah, Lieh-Ting, Yinglei, Joshua and Pragya for

their support and company.

My thanks also go to friends outside of Kan research group. I thank Dong Yang from Molnar group for his help performing wire-bonding for my test chips, and Ivan Buckreyev for his tutorial on high performance voltage regulators. I thank all my friends for their friendship and all the fun we had during the past six years: Nan Xu, Yunye Gong, Mengjie Yu, Lingfeng Cheng, Chengyu Liu, Haibing Wu, Ying Niu, Shuning Jiang, Moyang Wang and Yi Jiang.

Last but not least, I would like to thank my parents Xiaoming Cao and Yinmei Hong, and my fianc` Christopher Torng for their unconditional love and support.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 CMOS process variation: origin and challenge	2
1.1.1 Effects of process variation on RF-DC rectifiers	4
1.1.2 Concepts of RF-DC rectifier based on tunable threshold transistors	6
1.2 Variability in biological systems	8
1.2.1 Effects of variability on enteric neural recording	9
1.2.2 Concepts of dynamic spike sorting based spike classification	10
1.3 Chapter organization	11
2 RF-to-DC rectifiers: applications and challenges	13
2.1 Introduction to RF energy harvesting	13
2.2 Review of prior works on RF-to-DC rectifiers	14
2.3 Operational Principles for the RF-to-DC rectifier	16
2.3.1 Modeling of Dickson RDR	16
2.3.2 Optimal threshold voltages in a single stage of the RDR . .	18
3 Tunable threshold diodes: design and simulation	22
3.1 Structure of the tunable threshold diodes	22
3.2 SPICE simulation of the tunable- V_{th} diode	24
3.3 Simplified spice circuits for design optimization	27
3.4 Experimental measurements of the tunable- V_{th} diodes	28
4 Implementation and optimization of the rectifier system	33
4.1 Circuit implementation and experimental methods	33
4.2 Algorithm for the V_{th} optimization	36
4.2.1 Algorithm description	36
4.2.2 Automatic optimization in Cadence simulation	38
4.2.3 Optimization results for the tunable RD-to-DC rectifier . .	40
5 Experimental results of the tunable threshold rectifier	43
5.1 Sensitivity and efficiency	43
5.2 Retention time	49
5.3 Process variation tolerance	49
5.4 Comparison with prior works	50
5.5 Conclusion	52

6	Exploration of the tunable diodes used in cross-coupled rectifiers	54
6.1	Design of cross-coupled rectifiers using tunable diodes	54
6.2	Performance comparison	57
6.3	Summary	60
7	Variability in Enteric Neural Recording: Source and Resolutions	61
7.1	Introduction to the Enteric nervous system and its variability . . .	61
7.2	Prior works in spike classification for Enteric neural recording . .	62
7.3	Proposed new classification method for neural recording with large variability	64
8	Fast dynamic time warping spike classification algorithm: method and performance	65
8.1	introduction	65
8.2	fastDTW algorithm	65
8.2.1	DTW in similarity calculation	65
8.2.2	Overview of the fastDTW method with automatic thresh- olding	69
8.3	Performance analysis	73
8.3.1	Accuracy analysis	74
8.3.2	Computational complexity	83
9	Experimental spike classification by fastDTW	86
9.1	Experimental methods	86
9.2	Variable spike waveforms of experiment data	88
9.3	Discission on the nonstationary effects	91
9.4	Conclusion	94
10	Conclusion	95
10.1	Summary of major contributions	95
10.1.1	Contributions in RF-to-DC rectifier design	95
10.1.2	Contributions to enteric recording in high noise environ- ment	96
10.2	Suggestions for future work	97
10.2.1	Future works on the RF smart sensing platform	97
10.2.2	Future works on the spike sorting algorithm	100
	Bibliography	102

LIST OF TABLES

4.1	Input impedance of the rectifiers	35
5.1	Summary Of Rectifier Performance and Comparison with Prior Arts	53
6.1	Parameters for the conventional and the tunable cross-coupled rectifiers	56
8.1	Complexity per spike of different methods.	85
8.2	Variables to estimate the computational complexity.	85
9.1	Correlation between the spike magnitude and half-width for mechanical and chemical stimulation.	93

LIST OF FIGURES

2.1	(a) Schematics of the Dickson RF-to-DC rectifier. (b) A diode-connected transistor is typically used as the diode in each stage. We further implement the floating gate structure to achieve tunable- V_{th}	17
2.2	Effects of the offset voltage and load on a single stage in the RDR. (a-d) Left plots depict the leakage current $I_{leak,1}$ and the ratio of the positive current and the leakage current $\frac{I_{pos,1}}{I_{leak,1}}$ as a function of the threshold voltages $V_{th,0} = V_{th,1} = V_{th}$ for the cases when (a) $V_{offset} = 0, I_{load} = 0$, (b) $V_{offset} = 0, I_{load} = 2\mu A$, (c) $V_{offset} = 0.5V, I_{load} = 0$, (d) $V_{offset} = 0.5V, I_{load} = 2\mu A$, respectively. (a-d) right show how the output voltage of the single stage varies with the threshold voltages of the input diode $V_{th,0}$ and the output diode $V_{th,1}$ of the stage. Here the cases with $V_{offset} = 0$ mimic the 1st stage, while those with $V_{offset} = 0.5V$ mimic the last few stages of the RDR.	21
3.1	Schematics of the tunable diode in different operation modes. Circuit connection and band diagram of the tunable diode in (a) program mode; (b) erase mode. (c) Illustration of the tunable diode in the charge pump for RF-DC conversion.	23
3.2	Schematic of the SPICE circuit model for the tunable- V_{th} diode with Verilog-AMS programmed gate-oxide tunneling current source.	25
3.3	A summary of the gate-oxide tunneling current in a MOS transistor in different operational regions by the BSIM4 model. . . .	26
3.4	The simulated gate-oxide tunneling current, and its constituent components as a function of the gate voltage.	26
3.5	The transient simulation of the tunable- V_{th} diode by the proposed SPICE model when the program voltage is -3.45 V. A V_{th} shift of about 610 mV was achieved.	27
3.6	Schematics of the tunable diode with a manually added floating gate port for faster simulation.	28
3.7	Measured I-V curves of the 1.8 V tunable diode after program and erase operations (a) in linear scale (b) in log scale.	29
3.8	Measured I-V curves of the 3.3 V tunable diode after program and erase operations (a) in linear scale (b) in log scale.	31
3.9	Retention experiments of the tunable- V_{th} diodes after the V_{th} has been reduced by erase operations for (a) the 1.8 V diode (b) and the 3.3 V diode.	32
4.1	The block diagram of the control circuitry for different operational modes.	34

4.2	(a) The die picture of the fabricated chip. The zoom-in view shows the layout of the 8-stage tunable RDR together with its control circuits. (b) Experimental setup including the custom PCB to interface with MCU and the SMU. (c) System diagram of the experimental setup. The green dashed outline and the grey box indicate PCB level and chip level circuits.	34
4.3	Algorithm to find the optimal V_{th} s in RDR.	36
4.4	Flow chart for the algorithm for automatic V_{th} optimization of the tunable RF-to-DC rectifier.	37
4.5	The architecture for the simulation set-up for automatic optimization and data acquisition in Cadence Virtuoso simulations. .	39
4.6	Optimization of V_{th} for the maximum DC output voltage in (a) simulation and (b) experiment for the 1.8 V tunable- V_{th} RDR. (a) The left, middle and right figures demonstrate the results of the first, second and last round of optimization when $V_{in,peak} = 150mV$. (b) Experimental optimization process with $V_{in,peak} = 140mV$	40
4.7	The simulated optimal V_{th} values for the 1.8 V and the 3.3 V RDR with different operating conditions.	41
5.1	Performance comparison of the 1.8 V and 3.3 V 8-stage tunable- V_{th} RDR and the zero- V_{th} RDR, when the output load is open circuit. The solid lines and the dashed lines represent the results from the simulation and the experiments, respectively. The red, orange and blue lines represent the 1.8 V tunable- V_{th} , 3.3 V tunable- V_{th} and the zero- V_{th} rectifiers, respectively.	44
5.2	(a) Performances of the 8-stage tunable- V_{th} rectifiers under different optimization conditions and the zero- V_{th} RDR, as a function of the output current. (b) Retention test of the 1.8 V and 3.3 V tunable- V_{th} rectifiers optimized with $1\mu A$ output current. $V_{in,peak} = 140mV$ for both (a) and (b).	45
5.3	Simulation results of the 8-stage tunable RDR with $5000+821j$ input impedance and the $Q = 10$ matching network, when the output load is $500k\Omega$. (a) The peak voltage at the input of the matching network (red line) and the DC output voltage of the RDR (blue line) during the rectifier start-up, when the input power is -27 dBm (14 mV peak voltage). (b) The DC output voltage (red line with circular markers) and the PCE (blue line with diamond markers), as a function of the input sensitivity.	47
5.4	Distribution of the V_{out} of the tunable- V_{th} RDR with V_{th} variation. $V_{th,i}$ follows Gaussian distribution with ≈ 50 mV variance. The dark blue bars are the measured results, and the light blue shaded curve is the fitted distribution.	50

6.1	Schematic of a single stage for the cross-coupled rectifier. Cited from [5].	55
6.2	Simulated DC output voltage V_{out} as a function of $V_{in,peak}$ for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes.	57
6.3	Total PCE and its components as a function of input power, for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes. The solid line with circular markers, the dashed line and the dash-dot line represent the total PCE of the system, the PCE in the matching network and that in the rectifier.	58
6.4	Simulated DC output voltage V_{out} as a function of the input power for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes.	59
8.1	Similarity calculation using direct DTW. (a)(b): Template (blue solid lines) and candidate spike (blue dash dot lines) with the matched points connected by red dashed lines. (c)(d): Similarity matrices with the minimum-distance warping paths (red) for the spikes in (a) and (b).	66
8.2	FastDTW expedites direct DTW to linear time complexity. At each resolution stage, the optimal warping path is found in a window (colored square) containing the lower resolution warping path (red solid line).	68
8.3	Flow chart of fastDTW with automatic threshold decision. (a) Band-passed data of EAP. (b) Low-passed (100Hz) waveforms (blue) on top of the raw signal (grey) with the local maxima (red) detected as candidate spikes. (c) Magnitude distribution of candidate spikes, with a threshold determined as the local minimum to the right of the largest maximum. (d) Spike template used in classification. With no available templates (middle left), the estimated spike templates (red) are obtained by averaging the candidate spikes (grey) above the threshold in the first few time bins during initialization. (e) The similarities between the candidate spikes and the template are calculated by fastDTW. (f) The automatic similarity threshold is determined in the same way as in (c). Candidate spikes with larger similarity than the threshold are assigned to the template group. (g) A new averaged spike waveform is calculated from the classified spikes in the time bin, which can modify the previous spike template adaptively to account for the continuously slow change.	70
8.4	EAP waveforms of (a) the biphasic and (b) the monophasic spikes	75

8.5	Comparison of fastDTW and CCTM in similarity calculation: (a) Similarity calculated by fastDTW with spike misalignment (top) and duration variation (bottom). (b) Similarity calculated by CCTM with spike misalignment (top) and duration variation (bottom). BP stands for biphasic, and MP stands for monophasic. The grey lines represent the minimal similarity threshold that ensures no misclassification in the ideal case without noise. .	77
8.6	R_{TP} (a) and FP + FN (b) as a function of SNR by different methods. For fastDTW and CCTM, the similarity threshold is chosen to minimize the FP + FN. Spike duration $\propto \text{norm}(1,0.3)$ (Gaussian distribution) in the synthesized recording.	78
8.7	Classified spikes projected to the plane of the first two principle components, by PCA + k-means clustering with and without TW, at SNR=10 (a) and SNR=3 (b). Spike duration $\propto \text{norm}(1,0.6)$ in the synthesized recording.	79
8.8	ΔR_{TP} (red) and $\Delta \text{FP} + \text{FN}$ (blue), between biphasic and monophasic spikes, by different methods. Spike duration $\propto \text{norm}(1,0.3)$ in the synthesized recording.	81
8.9	R_{TP} (a) and FP + FN (b) as a function of relative similarity hard threshold by fastDTW and CCTM. Spike duration $\propto \text{norm}(1,0.3)$ in the synthesized recording.	82
8.10	Trade-off between the FP+FN rates (averaged over all SNRs) and CFOM for different methods. Spike duration $\propto \text{norm}(1,0.3)$ in the synthesized recording.	84
9.1	(a) Morphology of the neurons in mouse ENS. (b) A two-compartment circuit model of the neuron and CvMOS sensor interface. (c) A picture of the CvMOS sensor (the small chip on glass), and the electrode chip (the large chip with the fluidic well) used for the EAP recording.	87
9.2	Bandpassed experimental recording (1st row), firing rate of spikes classified by fastDTW (2nd row) and by CCTM (3rd row), averaged spike magnitude (4th row) over time for (a) mechanical stimulation, (b) chemical stimulation, and (c) TTX inhibition. (d) A zoom-in view of the 0.5-second segment under the mechanical stimulation, showing the spike train. Notice that the waveform features with magnitude larger than 3mV under mechanical stimulation around 40 s were not recognized as action potentials, but more likely due to movement between tissues and electrodes. Due to the large variability in ENS EAP recording, the robust classification algorithm is very critical, because user inspection can be misleading.	90

9.3	(a) The Box-and-whisker plot for the half-widths of the mechanically induced spikes; The aligned waveforms of (b) MP and (c) BP spikes before, during and after the mechanical stimulation. The red lines represent the medium waveforms. The nonstationary waveform features can be clearly observed.	91
9.4	(a) The Box-and-Whisker plot for the half-width of the chemically induced spikes. The averaged waveforms over time, for the MP spikes (b) and BP spikes (c).	92
9.5	The Box-and-whisker plots of spike half-width and magnitude vs. the firing rate, (a) for mechanically induced spikes, and (b) for chemically induced spikes.	92

CHAPTER 1

INTRODUCTION

Randomness and variation are ubiquitous in the physical world. Two thousands years ago, the great Greek philosopher Aristotle started to notice randomness in nature and classified events into certain, probable and unknowable ones [33]. Another philosopher Epicurus believed that randomness is independent of human knowledge [92], and the randomness at atom level would bring about uncertainties at higher levels. Since the ancient philosophers brought forward the idea of randomness [4], almost all science and engineering fields have witnessed such facts. In 1600s, mathematicians noticed the presence of stochasticity in digit sequences of square roots, logarithms, and numbers like Π [117]. In 1803, Thomas Young discovered that a single photon passing through double slit exhibits non-deterministic patterns in its path [14], shedding light onto the wave-particle duality of light. Around the same era, researchers observed the phenomenon of unpredictable swirls in liquids and air, leading to the theory of fluid turbulence [15] today.

Modern complementary metal oxide semiconductor (CMOS) technology is no exception from the randomness and variation, especially when the dimension of the functional device scales down to nano-meter regime where randomness takes the form of process variation originated from limited number of dopants and even atoms. Variation is also a built-in feature in biological systems, manifested in both large ecological systems and small microbiological units.

In this dissertation we present possible methods to deal with the variabilities in CMOS sensing networks. On one hand, the variation can originate from

the sensor tag itself, where the performance can be compromised by the uncontrollable variations. We will use the process variation effects on the RF-to-DC rectifiers with process variation as an illustration, as many passive sensors need such units to scavenge the ambient energy to accomplish the sensing functions. Process variation will reduce the sensor tag sensitivity and distort the transduction accuracy. We will present tunable device features to compensate these device variations. On the other hand, the sensing variation can originate from the targeted biological sensing signal, which complicates both the sensor system design and the associated signal analysis. We will illustrate a spike-sorting method to reliably classify the enteric neural signals which have unique waveform features but large variation in magnitude, timing and duration. Other variation sources can also affect the sensor system design, but our approaches of device compensation based on operational feedback and signal tolerance based on time warping are able to give illustrations for sensor designers to successfully countermeasure uncontrollable variation sources.

1.1 CMOS process variation: origin and challenge

Continuous scaling in CMOS technologies has improved the performance and reduced the unit cost of silicon integrated circuits (IC) by orders of magnitude in the past decades [110, 9]. The reduced device feature size has increased the transistor count from thousands to billions and improved the microprocessor operating frequency from MHz to GHz [9]. The resulting exponential growth in memory capacity [85] and computation power has enabled the recent development of revolutionary applications like artificial intelligence[106, 49] and big data[65]. The decrease in the supply voltage and the threshold voltage from

scaling makes ultra low-power systems possible for Internet of Things[88, 17], wearables[12] and bio-medical applications[72]. To integrate the electronic systems further into the physical world that the user lives in and cares about, we have envisioned billions of sensors in the sensor networks from around the user body to the entire Internet.

The advancement brought by scaling also comes with new challenges in design and analysis – variations, as one of the most serious and unavoidable consequences. A large number of variation effects in CMOS technology have been reported and investigated, including but not limited to random dopant fluctuation[69], line roughness[27], local oxide thickness variations[6], interface charge non-uniformities[11] and patterning proximity effects[120]. As the transistor feature size decreases to fundamental dimensions, such variations become critical for IC performance[54]. For example, process variations can increase functional failures in SRAM, degrading memory yield[73, 38], and increase timing violations in modern microprocessors[22].

Accommodation of process variations becomes a central issue in the reliable design methodology in most functional units of processors, memories and sensors. As digital signals can employ signal regeneration and error correction, we opt to investigate the variation effects in mixed-signal systems and the possible resolution by compensation. Specifically in this thesis, we will focus on the impact of CMOS threshold-voltage variation on RF-DC rectifiers, which is a key component for RF energy harvesting systems of most passive sensors.

1.1.1 Effects of process variation on RF-DC rectifiers

The RF-to-DC rectifier (RDR) is an important component in many systems relying on energy harvesting including passive RFID (radio frequency identification) tags [10, 67], bio-medical implants[68], and smart sensor networks[71], where item identity (ID), ambient condition and wearer vital signals [41] can be wirelessly collected. Among all sorts of ambient sources such as thermal energy, light and mechanical vibrations, RF energy harvesting is advantageous due to its on-demand availability. In an RF energy harvester, an antenna receives the incoming RF signal, an impedance matching network maximizes the power transfer from the antenna to the rectifier, an RF-to-DC rectifier (RDR) converts the RF signal to DC voltage that powers the successive stages and the loads.

The power conversion efficiency (PCE) as two key metrics for RDR performance. The sensitivity sets the lowest level of the input RF signal that can turn on the rectifier. The PCE is the ratio of the output power to the load over the total input power. For applications where the system operates continuously and completely on harvested power, the rectifier design often focuses on the maximization of PCE and the output power with a reasonable sensitivity [53]. In other applications where the system has a wake-up circuitry and is not always on, the sensitivity is more crucial which determines the operating range [79]. The sensitivity is related to the turn-on voltage $V_{turn-on}$, defined as the minimum input voltage to the RDR to generate a usable output DC voltage in Eqs. (1.1) and (1.2); whereas PCE is limited by the diode dropout voltage V_{drop} .

$$\text{Sensitivity (dBm)} = 10 \log P_{in}(\text{mW}) \quad (1.1)$$

$$P_{in} = \frac{V_a^2}{2R_{ant}} = \frac{V_{turn-on}^2}{Q \cdot 2R_{ant}} \quad (1.2)$$

where V_a is the peak input voltage to the antenna, R_{ant} is the antenna impedance, and Q is the quality factor of the matching network.

The Dickson voltage doubler [21, 70, 98] is a popular CMOS implementation for RDR, where the transistor threshold voltage V_{th} sets the constraint for both $V_{turn-on}$ and V_{drop} . To improve the sensitivity and PCE, low- V_{th} or zero- V_{th} MOS transistors by strict fabrication control were used in [19, 51], which can bring $V_{turn-on}$ down to about 200 mV. However, strict V_{th} control to a nearly zero value is difficult, expensive and unavailable in many logic CMOS processes. Minute variation in the fabrication process as discussed in Sec. 1.1 can degrade the RDR performance significantly as the technology continues to scale. Furthermore, the leakage current increases significantly as V_{th} approaches zero, resulting in low PCE [58].

Fully cross-coupled (FX) rectifiers were utilized to overcome low PCE in [53], where V_{drop} is reduced significantly without increasing the leakage. FX rectifier with output feedback to the gate bias [35, 81, 5] can largely improve the dynamic range of the input power. Cross-coupled rectifiers utilize regular MOS transistors with about 500 mV V_{th} instead of special zero- V_{th} MOS transistors, and are thus less susceptible to process variations. However $V_{turn-on}$ remains high for FX rectifiers, which restricts the sensitivity.

To further reduce $V_{turn-on}$ without using special zero- V_{th} transistors, V_{th} compensation structures in Dickson rectifiers were proposed. A semi-passive external compensator was used in [112], but the requirement on battery poses additional restriction to the tag deployment and packaging. In passive systems, two

auxiliary rectifiers [30] were utilized for V_{th} cancellation, together with high- Q passive input voltage boost in the matching network to lower the sensitivity to -19 dBm. An adaptive forward and backward V_{th} compensation scheme [36] employed auxiliary transistors to reduce V_{th} in the positive cycle and to reduce leakage in the negative cycle. However the auxiliary circuits rely on the starting condition and add complexity and instability to the system. A floating-gate transistor based rectifier was designed in [55], which separates the V_{th} compensation operation from rectification, but the rectifier still suffers from high leakage with low V_{th} .

1.1.2 Concepts of RF-DC rectifier based on tunable threshold transistors

To overcome the impact of process variation and to solve other issues discussed previously, we propose a new Dickson RDR based on tunable- V_{th} diodes, where each diode can be individually programmed at the wafer testing stage for either low V_{th} or low leakage for best overall performance. The proposed RDR can effectively compensate process variation which enables adoption of general logic CMOS processes [66] without low- V_{th} or zero- V_{th} transistor options. Moreover, it can be readily used in combination with additional techniques such as high- Q impedance matching [55, 64, 103, 84] and tandem stages [10] for best sensitivity and PCE [1]. With a $Q = 10$ matching network, -27 dBm sensitivity and 22% efficiency can be achieved for about 0.5 V DC output to a 500 $k\Omega$ load at 570 MHz.

In the proposed tunable- V_{th} RDR, the diode-connected transistors in typical

Dickson rectifiers are replaced by tunable diodes. The tunable diodes are based on the single-poly embedded Flash memory structure [80, 66] consisting of one transistor and three capacitors with same oxide thickness and different area ratio. The device structure can be implemented in most standard logic CMOS processes, and is commonly used as the nonvolatile memory cell in RFID chips [10]. Two of the three capacitances are used to tune the effective V_{th} of the transistors through program and erase (P/E), and the other one completes the diode structure. The capacitors and the transistor terminals are controlled by transmission-gate switches that are closed for P/E, and open for RDR operation.

The operation of the tunable- V_{th} RDR consists of two phases. In Phase 1, the diodes will be tuned to their optimal V_{th} values experimentally by an optimization algorithm implemented on MCU, which takes the output voltage as feedback at given input RF signal level and output load. In phase 2, the tunable- V_{th} RDR works in the same way as conventional rectifiers with only the RF signal source, when MCU and the DC power supplies are disconnected to be entirely passive for the energy harvesting operation. Phase 1 of the operation can be done in the burn-in wafer testing stage in the manufacturing process, and the optimal V_{th} is expected to last for the lifetime of the chip.

Because the V_{th} tuning is done experimentally post-fabrication, it can effectively compensate the process variation effects. The proposed RDR also separates the tuning phase and the rectification phase, making the RDR ready for system integration. Each diode in the rectifier can be optimized separately for low V_{th} or low leakage to achieve the best sensitivity with a high PCE. Due to such tunability, the proposed rectifier can offer good performance for a range of different input power levels and output loads.

1.2 Variability in biological systems

Biological systems are full of variabilities at the genetic, developmental, organismal, species, population, or ecologic/community levels [34]. Although such variabilities have shaped the world to be as diverse and beautiful as we have seen today, they also pose challenges to the sensor design and signal analysis of the intricate biological systems.

One example of such biological systems is the nervous system. Neural recording has been a heated area of research, because it sheds light onto the fundamental question of neuroscience, how the neurons interact with each other and the outside world [29]. The neural recording results lay the basis for brain machine interface (BMI) applications, including deep brain stimulators for pain management and control of motor disorders, and vagal nerve stimulators for treating epilepsy[77, 57]. In the past decade, multiple non-invasive neural recording methods have been investigated in hope to replace or supplement the traditional intracellular method – patch clamp, which provides neural signals of the best quality at the risk of invasive surgical procedures[37, 50, 52]. The extracellular non-invasive recording is usually performed by an array of sensors in close proximity to the neuron network, which offers a wide field of sight together with high resolution[114, 74] for cellular and network levels of research. The neural spike variability is a concern for extracellular recordings in which the sensors are away from the neurons, and thus subject to electrode drift, non-stationary waveforms [7] and the noise from adjacent neuron groups. The spike variability makes neuron recognition and the functional study that follows difficult.

In this thesis, we will focus on the extracellular neural recording of enteric nervous system, which has high spike variability in magnitude, time and waveforms, and propose a new spike classification method with high variability tolerance.

1.2.1 Effects of variability on enteric neural recording

Enteric nervous system (ENS) is composed of 200-600 million neurons found in the gastrointestinal tract and plays a vital role in upholding gut functions of motility, epithelial secretion, and intestinal barrier [97, 28]. Disruption of ENS can cause Crohns disease, diabetes [118], irritable bowel disease (IBD) [105], and ulcerative colitis [8]. Different from those in the central nervous system (CNS), which is situated in the largely stationary brain, neural signals in the ENS are coupled to gut motility and peristalsis, interacting with gastrointestinal endocrine cells [107] and intestinal longitudinal muscles [111, 42]. These movements not only alter neural activities, but also cause electrode drift during recordings. Activities of the ENS is also influenced by the immune system via cytokines and mast cell tryptase [20, 62]. Furthermore, the ENS consists of various types of enteric neurons embedded in dense mesh-like 2-D ganglia, called myenteric and submucosal plexuses, and hence an extracellular electrode will pick up signals from an ensemble of heterogeneous neurons with different action potential waveforms. These factors, together with those discussed in the previous section, render ENS recordings large waveform variability, which is compounded by the lack of knowledge of the ENS. To reliably interpreting the ENS extracellular action potential (EAP) recording, a spike classification method with high variability tolerance is needed.

1.2.2 Concepts of dynamic spike sorting based spike classification

To deal with the high spike variability in enteric nervous system, we proposed a new spike classification algorithm based on dynamic time warping (DTW). The spike classification method consists of three steps: (1) detection of candidate spikes, (2) waveform feature extraction, and (3) clustering of spikes. In the proposed method, we used the spike waveform as the feature and introduced DTW as a similarity measure, which has proven to be very effective in speech recognition under source and ambient variations [122] and ECG profile characterization[40].

DTW greatly improves the tolerance of spike variabilities, but has quadrature complexity. Therefore we implemented an approximate algorithm for DTW with linear complexity using adaptive temporal gridding [96] with linear complexity. With the improved similarity measure, we established a fastDTW method with automatic thresholding, which is suitable for unsupervised real-time closed-loop applications.

The proposed fastDTW spike classification achieves remarkable enhancement in accuracy and computational complexity in comparison to cross-correlation based template matching and PCA + k-means clustering without time warping, and was applied on *in vivo* mouse enteric neural recordings when the waveform variability is more than millisecond.

1.3 Chapter organization

This dissertation presents effective countermeasures for the variabilities in CMOS sensing networks, with specific examples on RF-to-DC rectifiers with process variation, and enteric neural classification with high spike variabilities. For RF-to-DC rectification, a new device structure together with an operational optimization algorithm was developed for better sensitivity and power conversion efficiency. For enteric neural recording, a novel on-line automatic algorithm based on dynamic time warping with high accuracy and linear complexity was introduced, with experimental verification on *in vitro* mouse enteric neural recording.

Chapters 2 to 6 discuss the design and implementation of the RF-to-DC rectifier with process variation. Chapter 2 introduces RF energy harvesting, summarizes prior works on rectifier design, and studies the effect of transistor threshold voltage on the rectifier performance in both theory and simulations.

Chapter 3 presents the structure and operation of the tunable- V_{th} diodes. The V_{th} tuning characteristics of the diode were verified by individual device measurements and represented in a SPICE model with gate-oxide tunneling current. A simplified analytical model was later introduced to speed up circuit simulation for design parameter optimization of multi-stage rectifiers.

Chapter 4 presents the system implementation of the tunable- V_{th} rectifier using the new diodes including both the integrated circuit chip and the custom PCB design, followed by a description of experimental setup and procedures. An algorithm to find the optimal V_{th} values for each diode with improved linear complexity is validated in Cadence simulations, and used for experimental

optimization of the rectifier with different output loads.

Chapter 5 discusses the experimental results of the optimized tunable- V_{th} rectifier in its sensitivity and PCE, retention time and process variation tolerance. The performance of the proposed rectifier is summarized and compared with prior start-of-arts.

Chapter 6 explores the design of cross-coupled rectifiers using tunable- V_{th} diodes to improve the sensitivity in high-PCE rectifiers. Different design options as well as trade-offs are discussed through simulation results.

Chapters 7, 8 and 9 describe the fastDTW spike classification algorithm for enteric neural recordings. Chapter 7 summarizes the prior works in extracellular spike classification and explains the basic concept behind the fastDTW algorithm.

Chapter 8 presents the procedure of the algorithm and investigates its accuracy and computational complexity, in comparison to the conventional cross-correlation based template matching and k-means clustering method with and without time warping.

Chapter 9 demonstrates the effectiveness of the fastDTW algorithm by applying it to experimental mouse enteric neural recordings, with different stimuli and inhibitions in the presence of high noise levels. Statistical analysis on the time evolution of spikes recognized by fastDTW under different conditions were performed for functional study of the enteric nervous system.

Chapter 10 concludes the dissertation with research achievements and suggestions for future work.

CHAPTER 2

RF-TO-DC RECTIFIERS: APPLICATIONS AND CHALLENGES

2.1 Introduction to RF energy harvesting

Ambient energy harvesting has attracted increasing attentions of researchers across the world, because of its potential ability to power battery-less devices with perpetual lifetime for the realization of the Internet of Things. Among all sorts of energy sources such as thermal energy, solar energy and mechanical vibrations, RF energy harvesting is advantageous due to its relieved installation constraints. RF energy harvesting can be used in many systems, including passive RFID (radio frequency identification) tags [10, 67], bio-medical implants[68], and smart sensor networks[71], where item recognition, ambient condition and occupant vital signals [41] can be wirelessly collected and processed. In an RF energy harvester, an antenna receives the incoming RF signal, an impedance matching network maximizes the power transfer from the antenna to the rectifier, a RF-to-DC rectifier (RDR) converts the RF signal to DC voltage that powers the successive stages and the loads.

The RF-to-DC rectifier is a critical component in the RF energy harvester, with the sensitivity and the power conversion efficiency(PCE) as two key metrics. The sensitivity sets the lowest level of the input RF signal that can turn on the rectifier. The PCE is the ratio of the output power to the load over the total input power. For applications where the system operates continuously and completely on harvested power, the rectifier design often focuses on the maximization of PCE and the output power with a reasonable sensitivity [53]. In other applications where the system has wake-up circuitry and is not always

on, the sensitivity proves to be more crucial which determines the operating range [79]. The sensitivity is related to the turn-on voltage $V_{turn-on}$, defined as the minimum input voltage to the rectifier to generate a usable output DC voltage in Eqs.(2.1)(2.2); whereas the PCE is limited by the diode dropout voltage V_{drop} .

$$\text{Sensitivity} = 10 \log P_{in}(\text{mW}) \quad (2.1)$$

$$P_{in} = \frac{V_a^2}{2R_{ant}} = \frac{V_{turn-on}^2}{Q \cdot 2R_{ant}} \quad (2.2)$$

where V_a is the peak input voltage to the antenna, R_{ant} is the antenna resistance, and Q is the quality factor of the matching network.

2.2 Review of prior works on RF-to-DC rectifiers

Dickson type voltage doubler [21, 70, 98] is a popular CMOS implementation for RDR, where the transistor threshold voltage V_{th} sets the constraint for both $V_{turn-on}$ and V_{drop} . To improve the sensitivity and the PCE, low- V_{th} or zero- V_{th} MOS transistors by strict fabrication control were used in [19, 51], which can bring $V_{turn-on}$ down to about 200 mV. However, strict V_{th} control to a nearly zero value is difficult, expensive and unavailable in many logic CMOS processes. Minute variation in the fabrication process can degrade the rectifier performance. Furthermore, the leakage current increases significantly as V_{th} approaches zero, resulting in low PCE [58].

Fully cross-coupled (FX) rectifiers were utilized to overcome the low PCE in

[53], which lowers V_{drop} without increasing the leakage. Combining FX rectifier with output feedback to the gate bias, [35, 81, 5] were able to largely improve the dynamic range of the input power. However $V_{turn-on}$ remains high for FX rectifiers, which restrict the sensitivity.

To further reduce $V_{turn-on}$, V_{th} compensation structures in Dickson rectifiers were proposed in recent works. A semi-passive external compensator was used in [112]. But the requirement on battery poses additional restriction to the tag deployment and packaging. In passive systems, [30] utilized two auxiliary rectifiers for V_{th} cancellation, together with high- Q passive input voltage boost in the matching network to lower the sensitivity to -19 dBm. [36] introduced an adaptive forward and backward threshold-voltage compensation scheme using auxiliary transistors to reduce V_{th} in the positive cycle and to reduce leakage in the negative cycle. However the auxiliary circuits relying on the starting condition adds complexity and instability to the system. A floating gate transistor based rectifier was designed in [55], which separates the V_{th} compensation operation from the rectification. But the rectifier still suffers from high leakage with low V_{th} .

To overcome the aforementioned problems, in this dissertation we propose a new Dickson RDR based on tunable- V_{th} diodes, where each diode can be individually programmed at the wafer testing stage for either low V_{th} or low leakage for best overall performance. The effect of V_{th} on RDR performance and the operating principles of the tunable diode are presented in Section 2.3. The circuit implementation and the experimental procedures are given in Section 4, followed by a discussion on the algorithm to find the optimal V_{th} values. The experimental results are given in Section 5. The proposed RDR can effec-

tively compensate process variation which enables adoption in general logic CMOS process [66] without low- V_{th} or zero- V_{th} transistor options. Moreover, it can readily be used in combination with additional techniques such as high- Q impedance matching [55, 64, 103, 84] and tandem stages [10] for best sensitivity and PCE [1].

2.3 Operational Principles for the RF-to-DC rectifier

Transistor threshold voltage V_{th} is crucial to the performance of the RDR. In this section, we first theoretically analyze the effect of V_{th} on the output voltage and the leakage current by modeling a Dickson type RDR, then investigate the optimal V_{th} for a single stage in different positions of the RDR through simulation. Finally the concept of a tunable- V_{th} diode is proposed to overcome the limitations of the rectifier designs.

2.3.1 Modeling of Dickson RDR

Figure 2.1(a) shows the schematics of an N -stage RDR[21], which consists of $2N$ diodes and $2N$ capacitors. A unit stage is highlighted in the lower-left corner. We followed the inductor-less RDR design in [21, 70, 98] and set all the diodes M_i and all the storage capacitors C_i to be the same.

An analytical model of the steady-state RF-to-DC conversion of a half-stage consisting of one diode M_i and one capacitor C_i was given in Eq. (2.3) by [119].

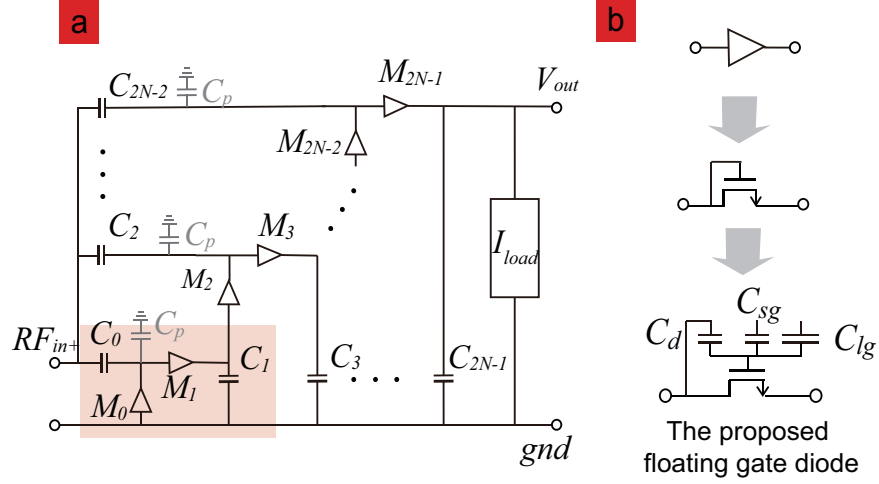


Figure 2.1: (a) Schematics of the Dickson RF-to-DC rectifier. (b) A diode-connected transistor is typically used as the diode in each stage. We further implement the floating gate structure to achieve tunable- V_{th} .

$$V_{out,i} = V_{in,peak} \frac{C_c}{C_c + C_p} - V_{th,i} - \left(\frac{15\pi I_{eff,i} \sqrt{2V_{in,peak}}}{8 \mu_n C_{ox} \frac{W}{L}} \right)^{\frac{2}{3}} \quad (2.3)$$

where $V_{in,peak}$ denotes the peak voltage and frequency of the input RF signal; $V_{th,i}$ is the threshold voltage of the i^{th} diode; $I_{eff,i}$ is the effective current through the diode M_i . C_c is the value of the stage capacitances, eg. $C_i = C_c$ for $i = 1, 2, \dots, 2N - 1$; C_p is the parasitic capacitance. Also μ_n denotes the electron mobility, C_{ox} the unit-area gate capacitance, and W and L the channel width and length of the diode-connected transistor, respectively.

From Eq.(2.3), we can see that to the first order, decreasing V_{th} can increase the DC output voltage, and therefor improve the sensitivity and PCE. However, as the V_{th} approaches zero, the leakage current starts to increase exponentially, which degrades the PCE drastically. The increasing leakage also reduces the output voltage through $I_{eff,i}$ in the 3rd term in Eq. (2.3), which is negligible for large V_{th} values.

The effective current $I_{eff,i}$ depends on both the load current and the leakage

current:

$$I_{eff,i} = I_{out,i} + \frac{I_{leak,i}}{\pi} \frac{W}{L} (1 - e^{-\frac{V_{in,peak}}{V_T}}) (1 + \lambda_{sub} V_{in,peak}) \quad (2.4)$$

where $I_{out,i}$ is the DC output current to the following stages or the load for the final stage, $I_{leak,i}$ denotes the leakage current through the diode M_i , V_T is the thermal voltage, and λ_{sub} is the sub-threshold channel-length modulation parameter.

The leakage current $I_{leak,i}$ is given by Eq. (2.5):

$$I_{leak,i} = \mu_n \sqrt{\frac{q\epsilon_{Si}N_{ch}}{2\phi_s}} V_T^2 \cdot e^{\frac{-V_{th,i}-V_{off}}{\kappa V_T}} \quad (2.5)$$

where κ is the subthreshold swing factor, V_{off} is the subthreshold offset voltage, q is the electron charge, ϵ_{Si} is the silicon permittivity, N_{ch} is the doping concentration in the channel, and ϕ_s is the surface potential.

The output voltage of a multi-stage RDR is $V_{out} = \sum_{i=1}^{2N-1} V_{out,i}$, which reaches the maximum when the $V_{out,i}$ is maximized for each stage. In the previous discussion, we know that the optimal V_{th} for a half-stage is non-zero. In the following section, we will show that the diodes in each stage can have quite different optimal V_{th} s, due to different loads and input voltage offsets.

2.3.2 Optimal threshold voltages in a single stage of the RDR

In this section we investigate the optimal V_{th} s for a single stage consisting of two diodes and two capacitors (Fig. 2.1a left corner) in different positions of the RDR. As shown in Fig.2.1, the i -th stage is stacked onto the $(i-1)$ -th stage (Fig.2.1), and therefore experiences an input voltage offset $V_{offset,i} \approx V_{out,i-1}$. On the other hand, the i -th stage is also loaded by the input of the following $N-i$ stages as well as the output current I_{load} . Based on these observations, we can

use $V_{offset,i}$ and $I_{load,i}$ in the i -th stage to mimic a single stage. The threshold voltages of the input and the output diodes are denoted as $V_{th,0}$ and $V_{th,1}$, which are tuned to their optimal values in Eq. (2.3) to maximize V_{out} under various load conditions.

The pre-layout SPICE simulation of the representative single stage is depicted in Fig. 2.2. Figs.2.2(a,b) illustrate the case for the first stage with zero offset, while Figs.2.2(c,d)) correspond to the last few stages. The effect of the load current is illustrated by Fig.2.2(a,c) with $I_{load} = 0$ and Fig.2.2(b,d) with $I_{load} = 2\mu A$.

The effect of the threshold voltages on the current components can be seen on the left column of Fig. 2.2. Here $I_{pos,i}$ denotes the peak positive current through the i -th diode when it is on; $I_{leak,i}$ denotes the peak leakage current when the i -th diode is supposed to be off. At steady state the current components satisfy Eqs. (2.6)(2.7).

$$\int_0^{\frac{T}{2}} I_{pos,1} dt - \int_{\frac{T}{2}}^T I_{leak,1} dt = I_{load} T \quad (2.6)$$

$$\int_{\frac{T}{2}}^T I_{pos,0} dt = \int_0^{\frac{T}{2}} I_{leak,0} dt + \int_0^{\frac{T}{2}} I_{pos,1} dt \quad (2.7)$$

For all cases, the leakage current $I_{leak,1}$ increases significantly with decreasing V_{th} close to 0, which is consistent with Eq. (2.5). As V_{th} increases, the ratio $\frac{I_{pos,1}}{I_{leak,1}}$ first goes up as $I_{leak,1}$ is suppressed, and then drops down after a certain point where V_{th} is too high to turn on the diodes. When $V_{offset} = 0.5V$ is present, the ratio $\frac{I_{pos,1}}{I_{leak,1}}$ stays high for larger V_{th} .

The right column of the Figs. 2.2 shows the effect of V_{th} s on the output voltage. For the first stage without load (Fig. 2.2(a)), V_{out} is low when V_{th} is either close to 0 or close to 0.4 V, and the optimal $V_{th,0}$ and $V_{th,1}$ to achieve maximal V_{out}

are 0.2 V and 0.25 V, respectively. When the same stage is loaded by $2\mu\text{A}$ current (Fig. 2.2(b)), V_{out} becomes negative for bigger V_{th} where such load cannot be driven. The V_{out} profile shifts towards smaller V_{th} values to provide the extra current to load, with optimal $V_{th,0}$ and $V_{th,1}$ both at 0.1 V. For the last few stages without load (Fig. 2.2(c)), V_{out} is relatively flat for small V_{th} , and drops quickly as V_{th} approaches 0.4 V. The optimal $V_{th,0}$ and $V_{th,1}$ are both at 0.2 V. When a $2\mu\text{A}$ load is drawn from the last stage in Fig. 2.2(d), the optimal $V_{th,0}$ and $V_{th,1}$ shift to smaller values at 0.15 V.

In a multi-stage RDR, the earlier stage sees larger load currents due to leakage in the following stages, and smaller input offset. As a result, the earlier stages require smaller optimal V_{th} s than the latter. The distinct optimal V_{th} values call for a design in which each diode can be tuned to different threshold for the best performance, which is the motivation for our proposed tunable- V_{th} RDR. For a given load, the later stages see less stages that need to be further tuned, and thus reach the optimal states faster than the earlier stages. This trend is utilized in Sec. 4.2 for the automatic V_{th} tuning algorithm.

In this section, we have shown from the theory and simulation that the optimal V_{th} values for the diodes are non-zero, and can be different in different stages to achieve the best performance for the sensitivity and PCE. In the next chapter, we will demonstrate a novel diode structure with tunability in V_{th} , so as to overcome the effect of unavoidable process variation and to program each diode to its optimal operating point.

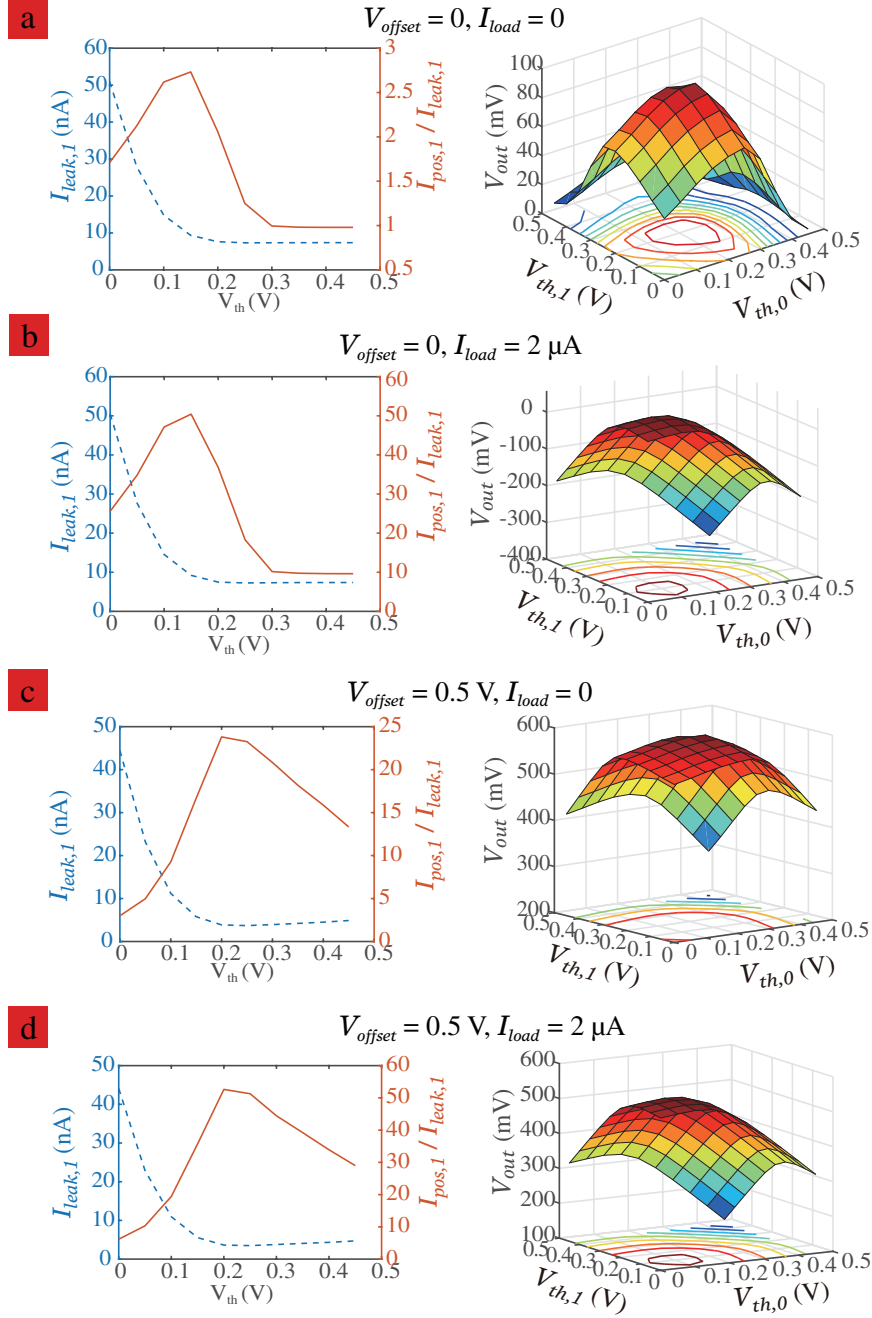


Figure 2.2: Effects of the offset voltage and load on a single stage in the RDR. (a-d) Left plots depict the leakage current $I_{leak,1}$, and the ratio of the positive current and the leakage current $\frac{I_{pos,1}}{I_{leak,1}}$ as a function of the threshold voltages $V_{th,0} = V_{th,1} = V_{th}$ for the cases when (a) $V_{offset} = 0, I_{load} = 0$, (b) $V_{offset} = 0, I_{load} = 2 \mu A$, (c) $V_{offset} = 0.5 V, I_{load} = 0$, (d) $V_{offset} = 0.5 V, I_{load} = 2 \mu A$, respectively. (a-d) right show how the output voltage of the single stage varies with the threshold voltages of the input diode $V_{th,0}$ and the output diode $V_{th,1}$ of the stage. Here the cases with $V_{offset} = 0$ mimic the 1st stage, while those with $V_{offset} = 0.5 V$ mimic the last few stages of the RDR.

CHAPTER 3

TUNABLE THRESHOLD DIODES: DESIGN AND SIMULATION

This chapter 3 presents the structure and operation of the tunable- V_{th} diode-connected transistors, which will be used in the proposed rectifier. The V_{th} tuning characteristics of the diode were verified by individual device measurements and represented in a SPICE model with gate-oxide tunneling current. A simplified analytical model was introduced for design parameter optimization of multi-stage rectifiers using the tunable- V_{th} diodes. The integration of the tunable- V_{th} diodes will be described in Chapter 4.

3.1 Structure of the tunable threshold diodes

To achieve tunable- V_{th} for individual transistor, we proposed a novel diode structure shown in Fig.2.1(b) bottom. The tunable diode is based on the single-poly embedded Flash memory structure [80, 66]. Compared to conventional diode-connected transistor, it has three extra capacitors C_d , C_{sg} and C_{lg} . The small-gate capacitor C_{sg} and the large-gate capacitor C_{lg} are used to tune the effective V_{th} of the transistors through program and erase (P/E); while C_d is to form the diode connection. The source node V_s , the drain node V_d , and the bottom plates of the capacitors C_{sg} and C_{lg} are controlled by transmission-gate switches that are closed for P/E, and open for RDR operation. The proposed device structure can be implemented in most standard logic CMOS processes, and is commonly used as the nonvolatile memory cell in RFID chips [10].

The program and erase (P/E) operations are enabled by a big capacitance ratio of $\frac{C_{lg}}{C_{sg}}$. In this work the area ratio of the capacitances is $C_{lg} : C_{sg} : C_d = 5 :$

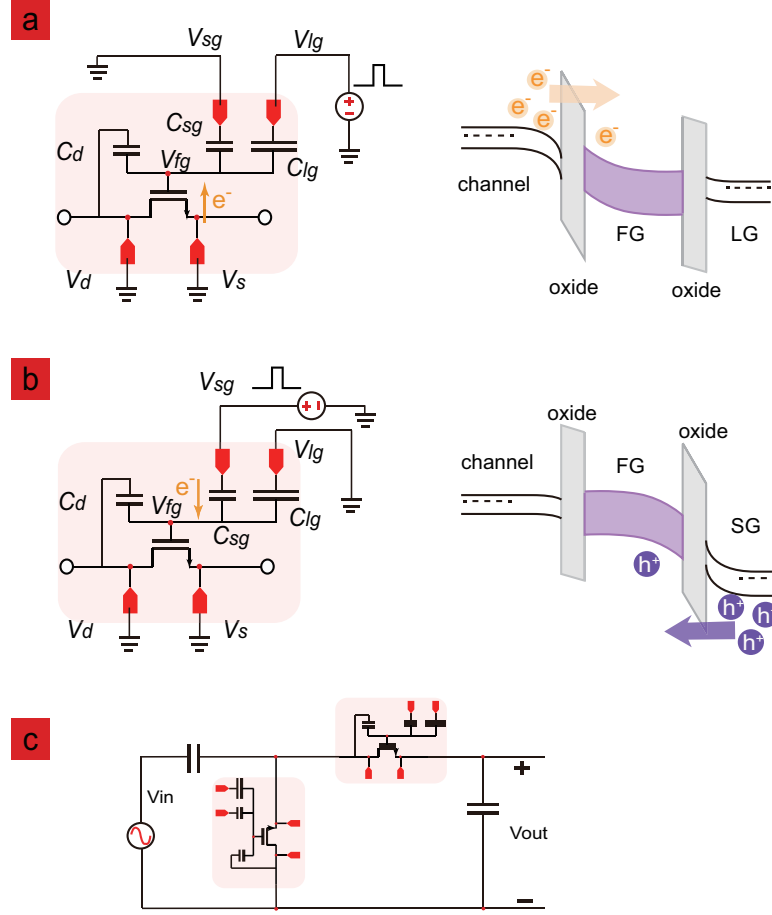


Figure 3.1: Schematics of the tunable diode in different operation modes. Circuit connection and band diagram of the tunable diode in (a) program mode; (b) erase mode. (c) Illustration of the tunable diode in the charge pump for RF-DC conversion.

0.4 : 1; the three capacitors are MOSCAPs of same oxide thickness. The floating gate voltage V_{fg} of the diode is expressed by Eq.(3.1):

$$V_{fg} = \frac{Q + V_{lg} \cdot C_{lg} + V_{sg} \cdot C_{sg} + V_d \cdot C_d + V_s \cdot C_{ox}}{C_{lg} + C_{sg} + C_d + C_{ox}} \quad (3.1)$$

where Q is the charge stored on the floating gate; C_{ox} is the gate capacitance of the transistor.

During program and erase, because the source and drain of the transistor are kept at ground, and that C_{lg} is significantly larger than other capacitances,

Eq.(3.1) reduces to $V_{fg} \approx V_{lg} + \frac{Q}{C_{tot}}$.

$$V_{fg} = \frac{Q + V_{lg} \cdot C_{lg}}{C_{lg} + C_{sg} + C_d + C_{ox}} \approx V_{lg} + \frac{Q}{C_{tot}} \quad (3.2)$$

During charge pump operation, all switches are open, C_{lg} (C_{sg}) in series with C_{ds} of the switches becomes negligible. Eq.(3.1) reduces to $V_{fg} \approx V_d + \frac{Q}{C_{tot}}$.

$$V_{fg} = \frac{Q + V_d \cdot C_d}{C_d + C_{ox}} \approx V_d + \frac{Q}{C_{tot}} \quad (3.3)$$

To turn on the transistor, V_{fg} must be larger than V_{th0} , which is the intrinsic threshold voltage of the transistor determined by the process. With charge on the floating gate, the effective threshold voltage of the diode V_{th} becomes $V_{th0} - \frac{Q}{C_{tot}}$.

The circuit connection and band diagram of the tunable diode in program mode is depicted in Fig.(3.1(a)). V_{lg} is connected to a high positive voltage, while V_{sg} , V_d and V_s are grounded. V_{fg} approximately equals to V_{lg} , which is high enough to induce electron injection from the channel of the transistor to the floating gate, which lowers the effective V_{th} of the diode. During erase operation (Fig.(3.1(b))), only V_{lg} is connected to a high positive voltage. V_{fg} approximately equals to V_{lg} which is ground. A high electric field appears across C_{sg} , inducing hole injection from SG to the floating gate, which increases the effective V_{th} . The total charge injected Q is a time integral of the Fowler-Nordheim Current I_{FN} by Eq.(13) in [91].

3.2 SPICE simulation of the tunable- V_{th} diode

In the previous section, we discussed the proposed structure of the tunable- V_{th} diode and the operation procedure for program and erase. To validate the

current, and its constituent components by the BSIM4 model, as a function of the gate voltage is depicted in Fig.3.4.

BSIM4 auxiliary function V_{aux} and parameters α , β and γ for gate tunneling current modeling				
Region of operation	Current component	V_{aux} , V_g	α, β, γ, p	
Acc.	J_{gb}	$V_{aux} = nighbacc \cdot V_t \ln \left[1 + \exp \left(-\frac{V_{gb} - V_{fbdb}}{nighbacc V_t} \right) \right]$ $V_{appl} = V_{gb}$	$\alpha = AIGBACC, \beta = BIGBACC, \gamma = CIGBACC, p = 1$	
Inv. and Depl.	J_{gb}	$V_{aux} = nigbinv \cdot V_t \ln \left[1 + \exp \left(\frac{V_{oxdepinv} - eigbinv}{nigbinv V_t} \right) \right]$ $V_{appl} = V_{gb}$	$\alpha = AIGBINV, \beta = BIGBINV, \gamma = CIGBINV4, p = 1$	
	J_{gc}	$V_{aux} = niggc \cdot V_t \ln \left[1 + \exp \left(\frac{V_{gsc} - V_{th}}{niggc V_t} \right) \right]$ $V_{appl} = V_{gsc}$	$\alpha = AIGC, \beta = BIGC, \gamma = BIGC, p = 1$	
All	J_{gs}	$V_{aux} = V_{gs} - V_{fbdb} $ $V_{appl} = V_{gs}$	$\alpha = AIGSD, \beta = BIGSD, \gamma = CIGSD, p = POXEDGE$	
All	J_{gd}	$V_{aux} = V_{gd} - V_{fbdb} $ $V_{appl} = V_{gd}$	$\alpha = AIGSD, \beta = BIGSD, \gamma = CIGSD, p = POXEDGE$	

Figure 3.3: A summary of the gate-oxide tunneling current in a MOS transistor in different operational regions by the BSIM4 model.

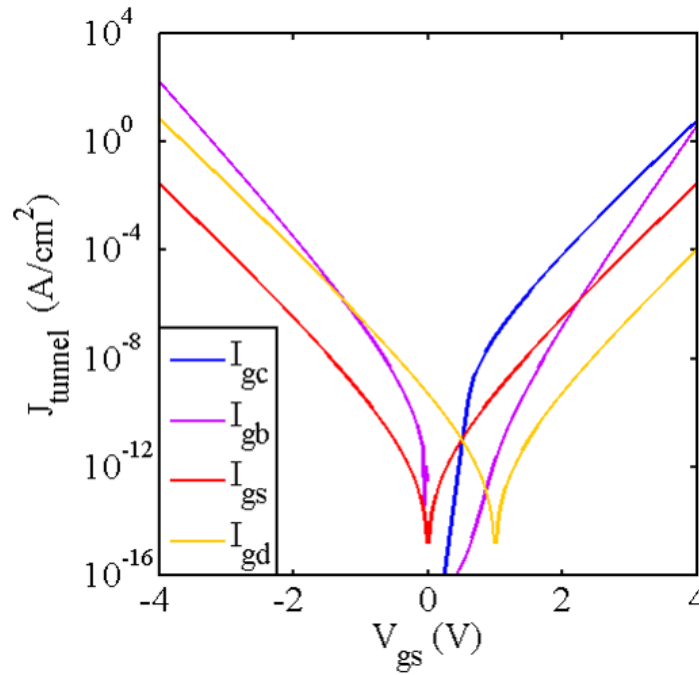


Figure 3.4: The simulated gate-oxide tunneling current, and its constituent components as a function of the gate voltage.

With the SPICE model described as above, we can verify the behavior of the tunable- V_{th} structure. An example of program operation in transient simulation is shown in Fig.3.5. It can be seen that the program voltage of -3.45 V induced

tunneling current which injects holes onto the floating gate, and a V_{th} shift of about 610 mV was achieved.

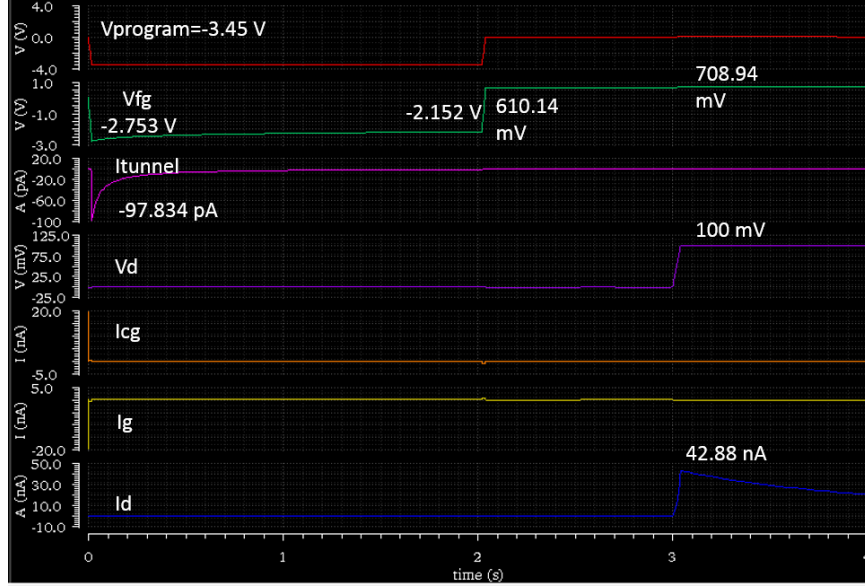


Figure 3.5: The transient simulation of the tunable- V_{th} diode by the proposed SPICE model when the program voltage is -3.45 V. A V_{th} shift of about 610 mV was achieved.

3.3 Simplified spice circuits for design optimization

In the last section, we have verified the functionality of the proposed tunable threshold diode qualitatively via Cadence simulation. However this simulation including both the P/E phase and the rectification phase is extremely computationally expensive. For example, one P/E operation and the following rectification stage of a single stage takes about 20 minutes. The simulation to program all diodes once in a 8-stage rectifier takes more than 3 hours to complete. In order to find the optimal threshold values for all stages, we would expect to perform tens of P/E operations on each diode, which can easily take years to

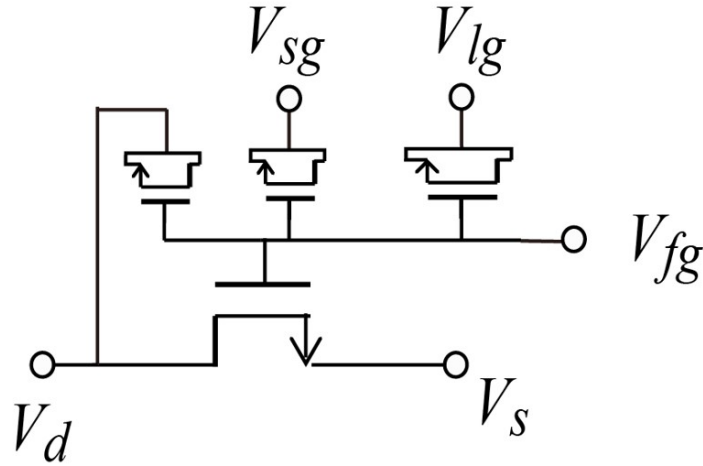


Figure 3.6: Schematics of the tunable diode with a manually added floating gate port for faster simulation.

finish.

In order to expedite the simulation process for rectifier design optimization, we modified the tunable diode structure virtually by adding an additional port on the floating gate (Fig.3.6). Through this port, we can set the shift in the threshold voltage ΔV_{th} for each diode using the "initial voltage" option in the Spectre simulator before the simulation. With such setting, the rectifier acts as if the assigned V_{th} values have been achieved by program and erase, and focuses on the different V_{out} and PCE performance for different threshold voltages. This simulation technique greatly simplified and expedited the simulation process, and was used in the rectifier optimization discussed in later chapters.

3.4 Experimental measurements of the tunable- V_{th} diodes

The proposed tunable- V_{th} diodes were designed and fabricated in a standard $0.18\ \mu m$ process. Program and erase operations were performed on individual

devices on the die using probe station to verify the tunability of the diodes.

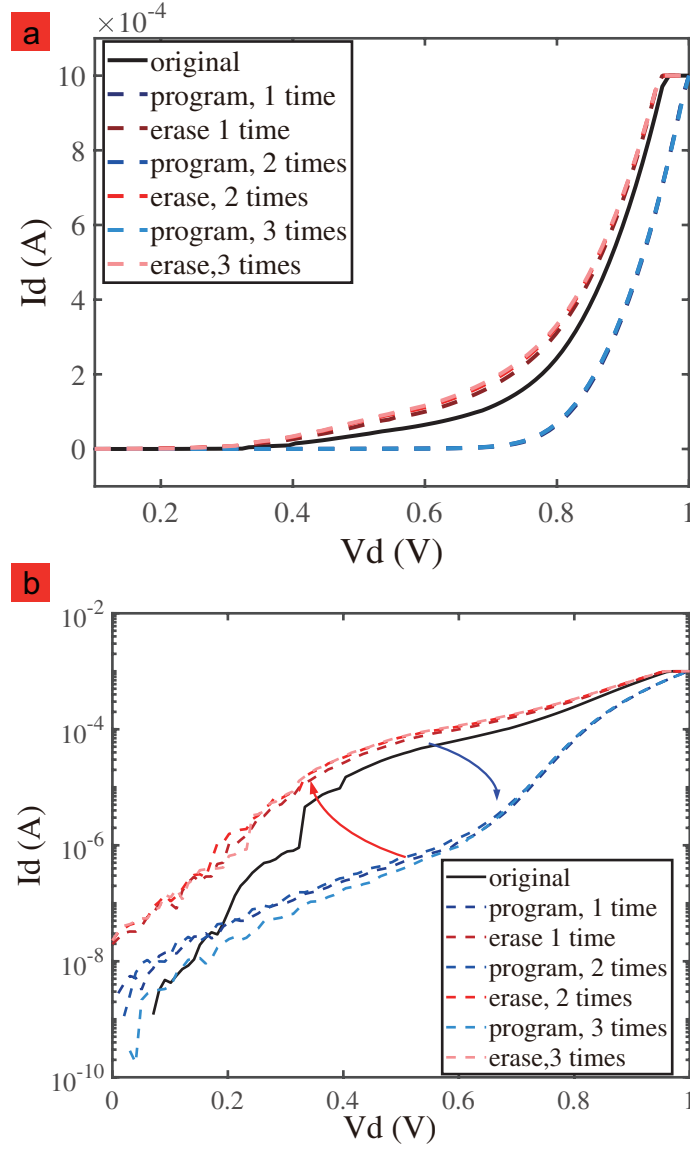


Figure 3.7: Measured I-V curves of the 1.8 V tunable diode after program and erase operations (a) in linear scale (b) in log scale.

Figure 3.8 shows the I-V curves of the 1.8 V tunable diode after several program and erase operations in linear scale and log scale. A voltage pulse of 3.3 V magnitude was used on either the small gate or the large gate of the diode structure for P/E as described in Section(3.1). It can be seen that program operations can increase the effective V_{th} of the diode, whereas erase decreases it. If we as-

sume a turn-on current of $1\mu A$, a tunability from 0.2 V to 0.7 V was achieved by program and erase.

The I-V curves of the 3.3 V tunable diode after several program and erase in Fig 3.8. A voltage pulse of 8.5 V magnitude was applied to induce P/E. Similar to the operation of the 1.8 V tunable diodes, program (erase) increases (decreases) the effective V_{th} of the 3.3 V diode. If we assume a turn-on current of $1\mu A$, a tunability from 0.1 V to 0.4 V was achieved by single program and erase operations.

Figure 3.9 demonstrates the results of retention experiments of the 1.8 V and 3.3 V tunable- V_{th} diodes. For the 1.8 V diode, the I-V curve starts to shift to the right towards its original position after 2 hours. After 23 hours, the threshold voltage has shifted back by half. This is because the 1.8 V diode has an oxide thickness of only 3.3-nm, and is subject to charge dissipation on the floating gate. For the 3.3 V diode with 6.5-nm oxide thickness, there's negligible shift in the I-V curve within 4 hours. After a retention time of 23 hours, the I-V curves shifts back very slightly.

Now that we've verified the functionality of the proposed tunable- V_{th} diodes both in SPICE simulation and in experimental measurements, we can move forward to the system design of the tunable- V_{th} RF-to-DC rectifier, which will be presented in details in Chapter 4.

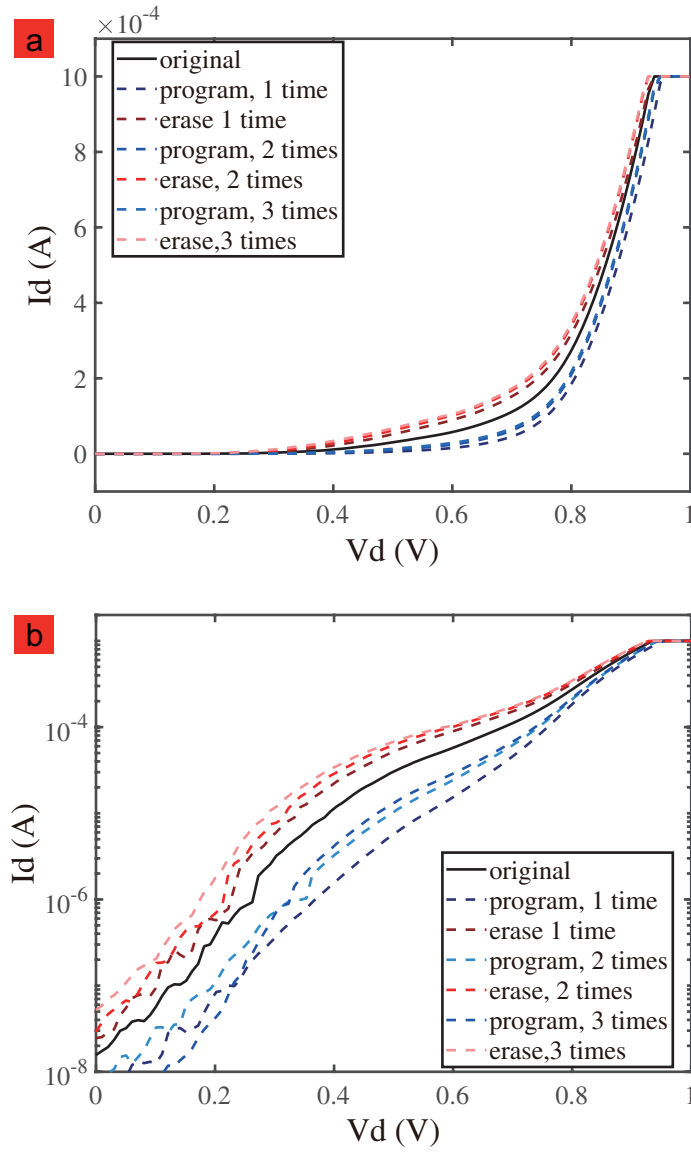


Figure 3.8: Measured I-V curves of the 3.3 V tunable diode after program and erase operations (a) in linear scale (b) in log scale.

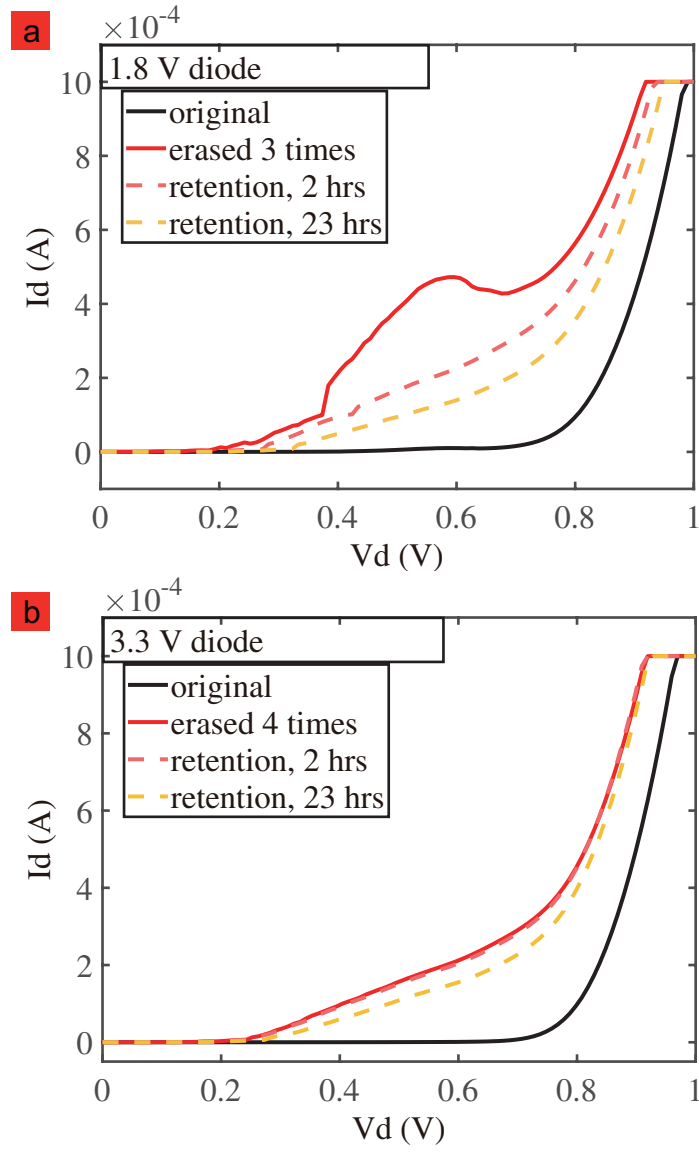


Figure 3.9: Retention experiments of the tunable- V_{th} diodes after the V_{th} has been reduced by erase operations for (a) the 1.8 V diode (b) and the 3.3 V diode.

CHAPTER 4

IMPLEMENTATION AND OPTIMIZATION OF THE RECTIFIER SYSTEM

In the previous chapter, we have introduced a novel diode structure with threshold voltage tunability. In this section we will discuss the system level implementation of the RF-to-DC rectifier using the tunable diodes. Section 4.1 demonstrates the circuit implementation for the rectifier, together with experimental set-up and procedures for the measurements. Section 4.2 establishes an automatic algorithm to find the optimal V_{th} values for all the diodes in different stages of the rectifier for the best performance. The algorithm is validated in simulation and the simulated optimal V_{th} values guide the experimental optimization and measurements.

4.1 Circuit implementation and experimental methods

The block diagram of a 8-stage tunable- V_{th} RDR is depicted in Fig. 4.1. As described in Sec. 3.1, diodes in RDR have switches connected to them that are open for RDR operation and closed for P/E operation. These switches form an array that controls individual diodes. The switches connected to the source and drain of the transistors are controlled by the enable signal $EN[0 : 15]$. The switches connected to C_{lg} and C_{sg} need to pass high voltage, and therefore are controlled by the level-shifted enable signal $EN_{ls}[0 : 15]$. The level shifter[99] translates 3.3 V signal to 5 V for the 1.8 V tunable- V_{th} RDR, and 5 V signal to 10 V for the 3.3 V RDR. The 16-bit EN and EN_{ls} signals are decoded from the 4-bit address generated by a microcontroller (MCU, Nucleo-F413ZH).

The die picture of the implemented chip in an UMC 0.18 μm process is shown

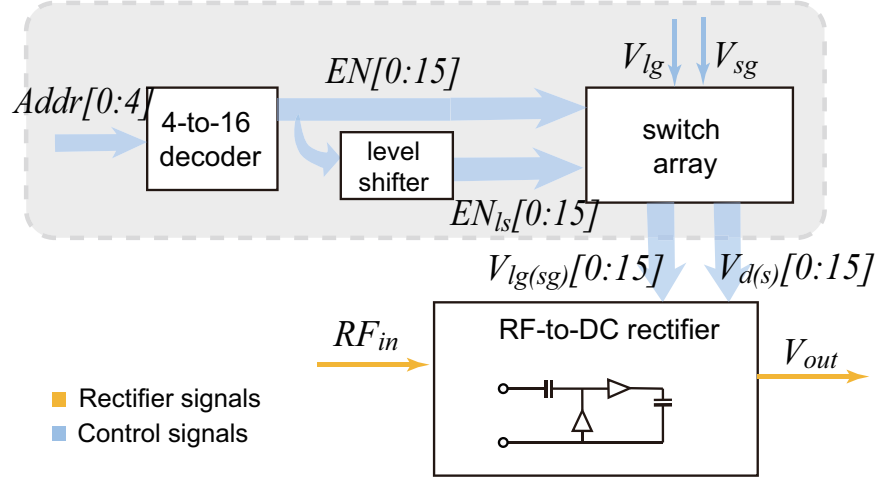


Figure 4.1: The block diagram of the control circuitry for different operational modes.

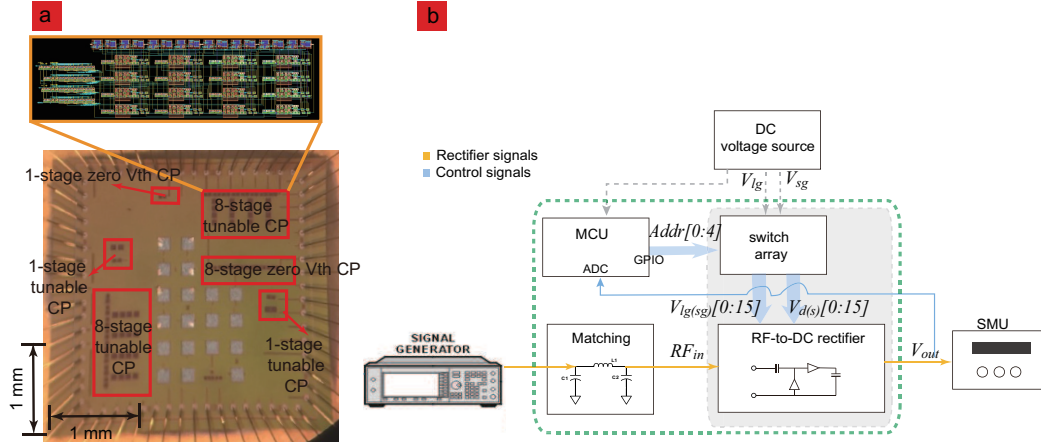


Figure 4.2: (a) The die picture of the fabricated chip. The zoom-in view shows the layout of the 8-stage tunable RDR together with its control circuits. (b) Experimental setup including the custom PCB to interface with MCU and the SMU. (c) System diagram of the experimental setup. The green dashed outliner and the grey box indicate PCB level and chip level circuits.

in Fig. 4.2(a). single stage and 8-stage tunable RDRs using 1.8 V and 3.3 V transistors, along with conventional zero- V_{th} RDRs were implemented for performance comparison. The zoom-in view shows the layout of the a 8-stage tunable RDR. For both the 1.8 V and the 3.3 V NMOS, the $\frac{W}{L}$ ratio is $9\mu m/360nm$.

Fig. 4.2(b) and (c) show the experimental setup for chip testing, where a

custom PCB provides the impedance matching network and other interface circuitry to the MCU and the signal sources. The operational procedure of the tunable- V_{th} RDR consists of two steps. (1) Optimizing V_{th} of each diode by the P/E algorithm described in Sec. 4.2. Pulsed square waves of 3.3 V (8.5 V) and 1 ms duration were used for P/E for the 1.8 V (3.3 V) transistors. In this step, both MCU and the RF signal source are on. The output voltage is monitored by the analog-to-DC converter (ADC) of the MCU. This step will be done in the burn-in wafer testing stage in the future manufacturing process, and the optimal V_{th} is expected to last for the lifetime of the chip. (2) Characterize the RF response of the RDR with only the RF signal source when MCU is disconnected. V_{out} with different loads is monitored by a source-measuring unit (SMU), and the entire unit does not have DC power supply to be entirely passive for the energy harvesting operation. The input impedance of the tunable- V_{th} RDRs and the zero- V_{th} RDR measured by the VNA are summarized in Table.4.1.

Table 4.1: Input impedance of the rectifiers

CP type		$R_p(\Omega)$	$C_{cp}(pF)$
8-stage	1.8 V tunable- V_{th}	85	13.4
	3.3 V tunable- V_{th}	107	35.6
	zero V_{th}	110	19.5
1-stage	1.8 V tunable- V_{th}	475	4.6
	3.3 V tunable- V_{th}	1120	6.2
	zero V_{th}	246	5.7

4.2 Algorithm for the V_{th} optimization

4.2.1 Algorithm description

Now that we have the ability to tune the diodes in the RDR individually, the next problem is to find the optimum V_{th} s for each stage. The most straightforward method one would think of is to perform a nested linear search for each diode, which costs $O(M^N)$ complexity, where M is the step number for each stage and N is the number of stages. If we assume a step number of 20 and 8 stages, the linear search requires 2.56^{10} attempts. With such complexity, it takes $> 10^5$ years to solve the optimum in simulation, which is impractical. The large number of P/E operations required in experiments also cause endurance issues of the chip.

Algorithm 1: Finding the optimal threshold voltages

```
initialize  $V_{out,max} = 0$ ; {The max  $V_{out}$  found in the search.}
for  $i = 1$  to  $number\_of\_stages$  do {rounds of optimization}
    initialize  $cur\_V_{out,max} = 0$ ; {The max  $V_{out}$  in this round}
    for  $j = 0$  to  $number\_of\_diodes - 1$  do {tune each diode}
        program( $j$ ); {program the  $j$ th diode}
        repeat
            erase( $j$ );
             $cur\_V_{out} = get\_v\_measurement()$ ;
        until a local max found or  $max\_erase\_num$  reached
    end for
    update( $cur\_V_{out,max}$ );
    if  $cur\_V_{out,max} \leq V_{out,max}$  then
        break; {search stops at convergence.}
    else
        update( $V_{out,max}$ );
    end if
end for
```

Figure 4.3: Algorithm to find the optimal V_{th} s in RDR.

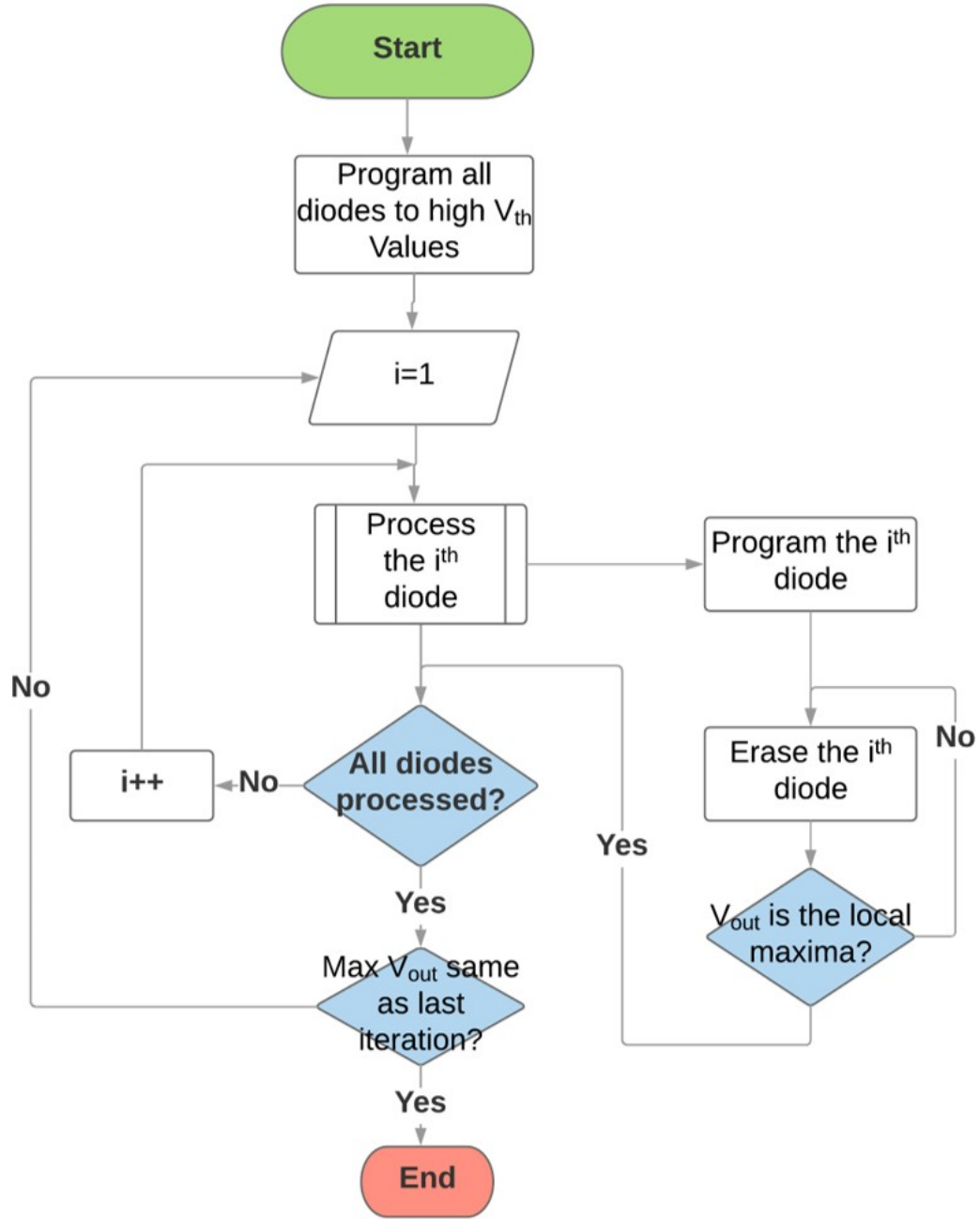


Figure 4.4: Flow chart for the algorithm for automatic V_{th} optimization of the tunable RF-to-DC rectifier.

To solve this problem, we developed an algorithm in Fig.4.3 that can find the optimum with $O(N)$ complexity. The initial V_{th} values are assumed to be sufficiently larger than zero. The algorithm first starts with erase operation to

bring V_{th} to lower values closer to optimum, and then search the surrounding V_{th} space iteratively to locate the optimal point. The convergence is achieved when two consecutive rounds of search give the same V_{out} . Based on the RDR performance observation in Sec. 2.3.2, the later stages take less iterations to converge. For example, the load to the last stage is completely set by the output, and the last stage is less sensitive to V_{th} variations. So it will achieve optimum in the first round. Once the operating point of the last stage is fixed, the load to the 2nd last stage is set and it will converge in the 2nd round. In the i -th round of the search, the diodes in the $N - i + 1$ -th stage would converge. Optimal V_{th} values can always be reached in N rounds. In experiments where the initial V_{th} values are unknown due to process variation, we can program the diode a few times to ensure V_{th} is larger than the optimum before running the algorithm. If the input RF signal level to optimize for is very low, e.g. $V_{in,peak} = 150mV$, the V_{out} changes caused by V_{th} shift may be too small to be detected by ADC to start the algorithm. In this case we can perform one round of search with higher $V_{in,peak}$ to set the initial V_{th} s closer to the optimum before performing the actual optimization. The flow chart of the proposed algorithm is depicted in Fig.4.4.

4.2.2 Automatic optimization in Cadence simulation

To verify that the proposed algorithm can tune the RF-to-DC rectifier to its optimum, we first perform the algorithm on a 8-stage RDR in Cadence Virtuoso simulation. Typically Cadence Virtuoso software offers a graphic user interface (GUI) for the users to access the simulation results. However this conventional method for data acquisition requires manual interactions of the users, including typing in the initial values of the circuits, clicking on the output waveforms and

manual data exportation, which is inconvenient for the optimization. Therefore we developed an automatic flow that can perform the simulation and data acquisition in an unsupervised manner.

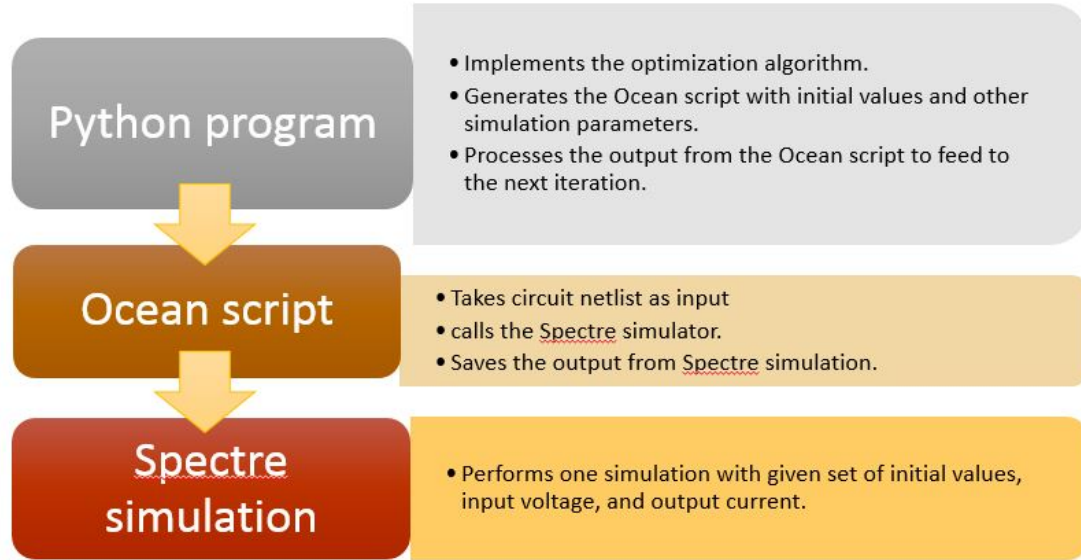


Figure 4.5: The architecture for the simulation set-up for automatic optimization and data acquisition in Cadence Virtuoso simulations.

The architecture of the developed simulation set-up is shown in Fig.4.5. The optimization algorithm is implemented in Python, which also generates the initial values and simulation parameters in an Ocean script. The Ocean script takes the netlist of the rectifier circuits as input, calls the Spectre simulator to perform simulation on the rectifier with one specific set of V_{th} values, and saves the output voltage to a given load in the designated format. The Python script reads in the simulation output and decides the next set of V_{th} values until an optimum has been found.

4.2.3 Optimization results for the tunable RD-to-DC rectifier

Figure 4.6(a) shows the simulated process to find the optimal V_{th} for each diode in the 1.8 V tunable- V_{th} RDR. In the simulation, the RF input signal has 150 mV peak voltage and 570 MHz frequency, and the output is open circuit. The left, middle and right figures demonstrate the search results of the first, second and the last rounds of the optimization algorithm. The initial values of the V_{th} s were purposely set to be far away from the optimum, and thus V_{out} was as low as 100 mV. The first round has brought each diode closer to its optimal V_{th} value, increasing V_{out} from 100 mV to about 650 mV. The diodes in the last few stages of the RDR, which are less sensitive to V_{th} variations than those in the front stages due to the established voltage offset, were almost tuned to their optimal states after the first round. Given the new initial condition from the result of the first round, the second round further optimized V_{th} of the front stages. As can be seen in the middle plot in Fig.4.6(a), V_{th} s of the first four diodes decreased from 300-500 mV to 50-100 mV. After the second round, most V_{th} were in close

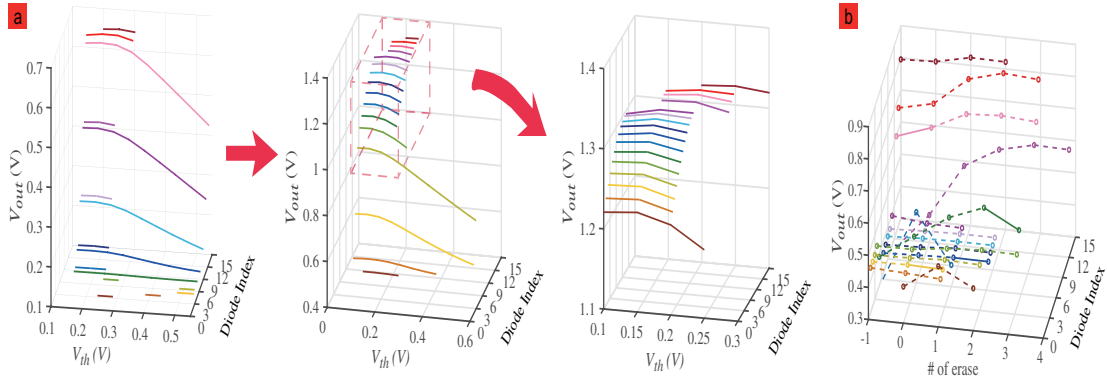


Figure 4.6: Optimization of V_{th} for the maximum DC output voltage in (a) simulation and (b) experiment for the 1.8 V tunable- V_{th} RDR. (a) The left, middle and right figures demonstrate the results of the first, second and last round of optimization when $V_{in,peak} = 150mV$. (b) Experimental optimization process with $V_{in,peak} = 140mV$

proximity of their optimal values, which were obtained in the third (last) round of the search. The search space of the last round was a subset of that of the second round, which was used as the convergence criterion.

The optimal V_{th} s found for the 0-11th diodes, the 12-14th diodes, and the 15th diode are 150, 200, 250 mV, respectively. For the front stages, lower V_{th} s are required so that the diode is sufficiently turned on for non-negligible I_{pos} . For the later stages, the leakage current has more impact on the output voltage, and a higher V_{th} to suppress the leakage can increase the output voltage.

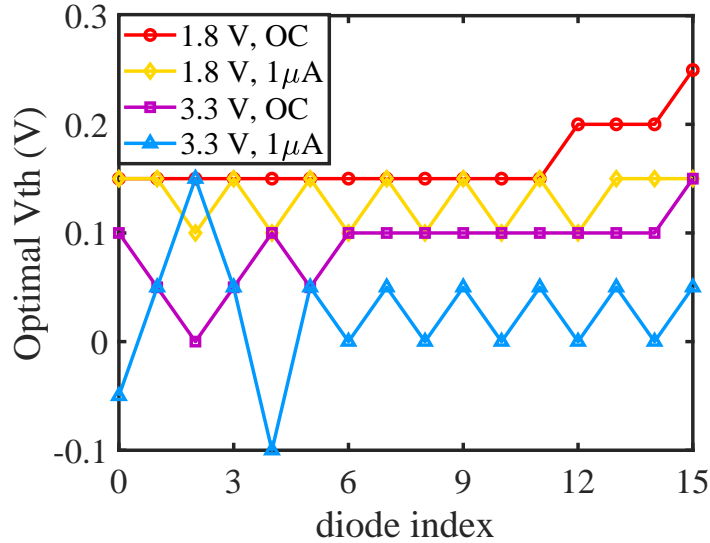


Figure 4.7: The simulated optimal V_{th} values for the 1.8 V and the 3.3 V RDR with different operating conditions.

Although the example demonstrated in Fig. 4.6 has an open circuit output, the optimization method works for cases with load as well. Fig. 4.7 depicts the optimized V_{th} values of the 1.8 V and the 3.3 V tunable- V_{th} RDR with different output loads during optimization. For the 1.8 V RDR with open circuit output (red line with circle markers), the optimal V_{th} value increases with the diode index. A similar increasing trend in V_{th} is observed for the 3.3 V RDR (purple line with square markers) too. This observation suggests that lower

V_{th} s should be used for the earlier stages to turn on the rectifier while higher V_{th} should be used for later stages to suppress the leakage. Overall the optimal V_{th} s of the 3.3 V RDR are smaller than those of the 1.8 V RDR, which is determined by the intrinsic properties of the devices. For the 1.8 V RDR loaded with $1\mu A$ current (yellow line with diamond markers), a zigzag pattern of the optimized V_{th} s can be seen. Within the same stage, the input diode (even-indexed) has lower optimal V_{th} than the output diode (odd-indexed), which implies the different functions of the input and output diodes. The 3.3 V RDR with $1\mu A$ output current (blue line with triangle markers) shows a similar pattern of V_{th} as the 1.8 V one, with smaller V_{th} s overall. For the 0^{th} and the 4^{th} diodes, the optimal V_{th} s are needed, which cannot be achieved by conventional compensation techniques [35, 5]. In contrast, the proposed tunable- V_{th} RDR can precisely control individual diode to achieve optimal and even negative V_{th} s for better performance.

CHAPTER 5

EXPERIMENTAL RESULTS OF THE TUNABLE THRESHOLD RECTIFIER

The fabricated tunable RF-to-DC rectifiers are optimized using the algorithm described in the previous section experimentally. In experiment, the algorithm is implemented in C/C++ on a MCU, which generates address information for the diodes and control signals for program and erase. During optimization, the output voltage is obtained by the MCU ADC for feedback. During the rectifier characterization, the output voltage is obtained from the SMU readout.

The experimental results of the proposed tunable rectifier, in comparison with a rectifier based on vendor provided zero-threshold diode are summarized in the following sub-sections. The rectifier performance is analyzed for the sensitivity, output voltage, PCE, retention time, and the tolerance to process variation. A comparison of the work with prior works can be found at the end of the chapter.

5.1 Sensitivity and efficiency

Figure 5.1 shows the resulting performances of the 8-stage 1.8 V and 3.3 V tunable- V_{th} RDR after optimization and that of the zero- V_{th} RDR. V_{out} is plotted as a function of the peak voltage of the input RF signal, when the output is an open circuit. The solid and dashed lines represent the post-layout simulation and experimental results, respectively. The red, orange and blue lines represent the 1.8 V and the 3.3 V tunable- V_{th} RDR and the zero- V_{th} RDR, respectively. The performance of the tunable- V_{th} RDRs are similar to those predicted by simulation, which indicates the optimum has been established during the P/E phase.

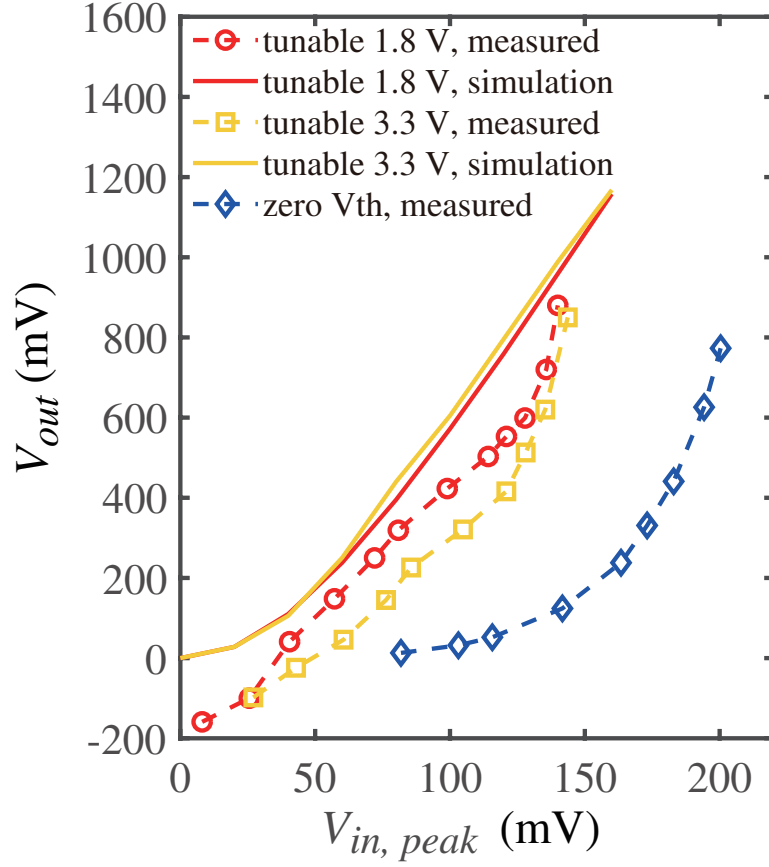


Figure 5.1: Performance comparison of the 1.8 V and 3.3 V 8-stage tunable- V_{th} RDR and the zero- V_{th} RDR, when the output load is open circuit. The solid lines and the dashed lines represent the results from the simulation and the experiments, respectively. The red, orange and blue lines represent the 1.8 V tunable- V_{th} , 3.3 V tunable- V_{th} and the zero- V_{th} rectifiers, respectively.

When the required V_{out} is 600 mV, both the 1.8 V and the 3.3 V tunable- V_{th} RDRs can work for input signal with $V_{in, peak}$ as low as 140 mV, in contrast to 200 mV for the zero- V_{th} RDR, corresponding to 3dB improvement in sensitivity. When the required V_{out} is 300 mV, which is often sufficient for many DC-DC converters to operate, the 1.8 V (3.3 V) tunable- V_{th} RDR works for $V_{in, peak}$ as low as 80 mV (100 mV), in contrast to 160 mV for the zero- V_{th} rectifier, corresponding to 6dB (4dB) improvement.

The 8-stage 1.8 V and 3.3 V tunable- V_{th} RDRs were optimized experimentally

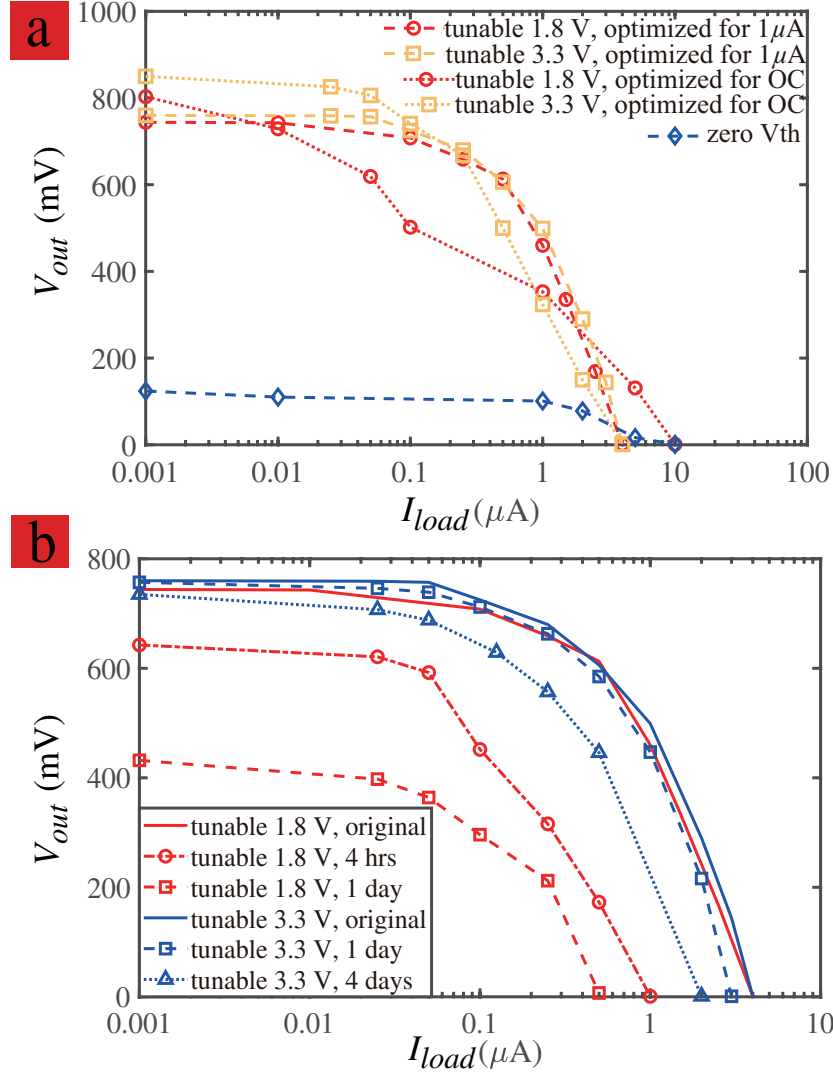


Figure 5.2: (a) Performances of the 8-stage tunable- V_{th} rectifiers under different optimization conditions and the zero- V_{th} RDR, as a function of the output current. (b) Retention test of the 1.8 V and 3.3 V tunable- V_{th} rectifiers optimized with $1\mu A$ output current. $V_{in,peak} = 140mV$ for both (a) and (b).

following the method in Sec.4.2 for performance comparison. An example of the experimental optimization of the 1.8 V RDR is illustrated in Fig.4.6(b). The general trend of the optimization path is similar to the prediction of the simulation, with some local variations due to the variation of the intrinsic V_{th} s of the diodes.

Figure 5.2(a) further illustrates how the output current I_{out} affects V_{out} for

different RDR implementations and optimization conditions when $V_{in,peak} = 140mV$. In all cases, V_{out} drops with increasing I_{out} . Both the 1.8 V and the 3.3 V tunable- V_{th} RDRs (red and orange lines) provide higher V_{out} than the zero- V_{th} RDR (blue line) in the entire I_{out} range up to $3\mu A$. The red dotted line shows the response of the 1.8 V tunable- V_{th} RDR with its V_{th} values optimized under the open circuit condition, which has indeed larger open circuit V_{out} than that optimized for $1\mu A$ load (the red dashed line). However, its V_{out} decreases faster with increasing I_{out} , which illustrates the important role of the V_{th} setting in RDR performance. Similar observations can be made for the 3.3 V RDRs optimized with open circuit (orange dotted line) and $1\mu A$ (orange dashed line). The 3.3 V RDR optimized with $1\mu A$ demonstrated as good performance as the 1.8 V RDR, verifying the usage of tunable- V_{th} RDR for general logic processes with different oxide thickness. When $I_{out} = 1\mu A$, both the tunable- V_{th} RDRs can generate $V_{out} = 500mV$, compared to 100 mV by the zero- V_{th} RDR (5 times or 14 dB improvement).

With the improved $V_{in,peak}$, the proposed tunable RDR can be used in combination with passive input voltage boosting in the matching network to further improve the sensitivity and the PCE. For example, when a matching network with $Q = 10$ is utilized, a sensitivity of -27 dBm (14 mV peak voltage) can be achieved. For maximal power transfer, the real part of the input impedance of the RDR needs to be increased to 5000Ω to match with the 50Ω antenna. The RDR and matching network co-design with such high input impedance is not straightforward, as the rectifier is a non-linear time variant system, and the input impedance changes during rectifier charging. [60, 2] designed the matching network by using the steady state input impedance. [56, 95] modeled the rectifier with a parallel RC circuit and found the optimum matching point through

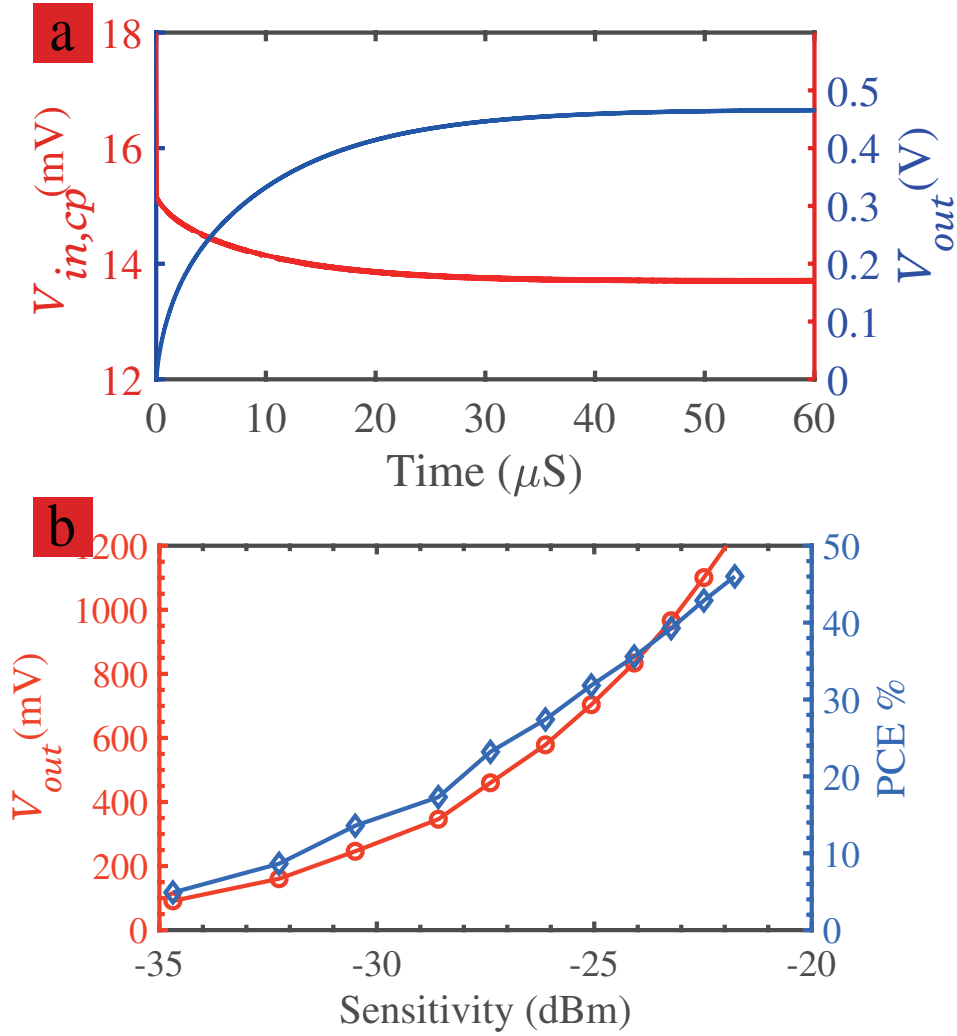


Figure 5.3: Simulation results of the 8-stage tunable RDR with $5000+821j$ input impedance and the $Q = 10$ matching network, when the output load is $500k\Omega$. (a) The peak voltage at the input of the matching network (red line) and the DC output voltage of the RDR (blue line) during the rectifier start-up, when the input power is -27 dBm (14 mV peak voltage). (b) The DC output voltage (red line with circular markers) and the PCE (blue line with diamond markers), as a function of the input sensitivity.

iterative simulations. [100] presented an analytical model for the transient input impedance of the rectifier during charging.

In this work, we designed a 8-stage tunable RDR with $5000+821j$ steady state input impedance together with a $Q = 10$ impedance matching network. The

dimension of the NMOS is $\frac{W}{L} = 350nm/500nm$. Fig.5.3(a) demonstrate the transient response of the designed rectifier, when the input power is -27 dBm and the output load is $500k\Omega$. The red line represents the peak voltage to the input of the matching network $V_{in,cp}$. The blue line denotes the DC output voltage. When good matching is achieved, $V_{in,cp}$ to the matching network is 14 mV, which happens after $10\mu S$ during charging. The input impedance is inversely proportional to the trans-conductance g_m of the diode, which is proportional to the V_{gs} of the transistor. When the RDR first turns on, V_{gs} is the highest and the input impedance is smaller than the steady state value. Therefore $V_{in,cp}$ at the matching network is higher than 14 mV. As the RDR continues charging, V_{out} increases gradually, bringing down V_{gs} .

The DC output voltage (red line with circular markers) and the PCE (blue line with diamond markers), as a function of the input sensitivity are plotted in Fig.5.3(b). Here the PCE is the total efficiency which includes the loss in the matching network and the rectifier efficiency. Both the output voltage and the PCE increase with sensitivity. When the input sensitivity is -27 dBm, a 22% total PCE was achieved with 460 mV output voltage to a $500k\Omega$ load.

Theoretically, rectifiers with even higher input impedance can be designed for even better sensitivity and PCE, if higher- Q components are available. But higher input impedance is usually obtained by decreasing the $\frac{W}{L}$ ratio of the transistor, which in turn reduces the charging current and therefore increases the start-up time of the rectifier.

5.2 Retention time

Retention experiments were carried out on the 1.8 V and 3.3 V tunable- V_{th} RDRs as shown in Fig.5.2(b). The 1.8 V and 3.3 V diodes have oxide thickness of 3.3 nm and 6.5 nm, respectively. The thicker the oxide, the longer the retention time is before P/E is needed to re-tune the V_{th} . Thicker oxide also means higher P/E voltage, which requires high-voltage IOs. For the 1.8 V RDR, both the open circuit V_{out} and the maximum I_{out} the rectifier can provide decreased noticeably after 1-day of retention due to the thin oxide. For the 3.3 V RDR, almost as good open circuit V_{out} and slightly smaller maximum I_{out} were observed after 4 days. Thicker oxide of 10nm is preferred for actual applications so that the rectifier only needs to be tuned once in house. But the experiments here serve as proofs of concepts for the proposed tunable- V_{th} RDR.

5.3 Process variation tolerance

Process variation is unavoidable in the fabrication and more and more serious as the technology continues to scale down. To study the impact of threshold variation on the output voltage, we performed Monte-Carlo simulation of the RDR. The RDR was first tuned to the maximal V_{out} by using the method described in Sec.4.2, and then additional P/E operations by MCU are performed to mimic the effect of V_{th} variation. For each data point, the number of P/E to be performed on the i -th diode follows a zero-mean unit-variance Gaussian distribution: $N_{erase,i} = \mathcal{N}(0, 1)$. Figure 5.4 shows the experimental distribution of the V_{out} with threshold variation. For Fig. 5.4, $V_{in,peak} = 140mV$, and the resulting maximal V_{out} is about 850 mV. The full width at half maximum (FWHM) of V_{out}

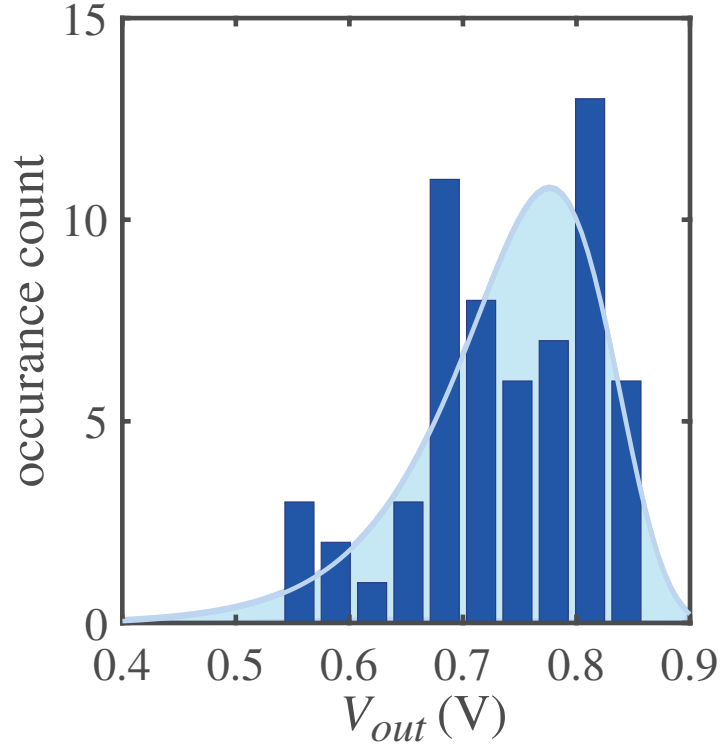


Figure 5.4: Distribution of the V_{out} of the tunable- V_{th} RDR with V_{th} variation. $V_{th,i}$ follows Gaussian distribution with ≈ 50 mV variance. The dark blue bars are the measured results, and the light blue shaded curve is the fitted distribution.

is 150 mV, indicating how much on average one-step (≈ 50 mV) of V_{th} variation can change V_{out} . If the V_{th} s of all diodes vary in the unfavorable direction, it has a big impact on the RDR performance, which can be overcome in our tunable- V_{th} RDR.

5.4 Comparison with prior works

The performance of the proposed RDR is summarized and compared with prior works in Table.5.1. The works in the top, middle and bottom parts of the Table are those based on Dickson type structure, cross-coupled architecture, and those

using additional techniques including very high- Q components and tending stages. Generally, the Dickson type architectures demonstrate a better sensitivity, whereas the cross-coupled designs provide higher PCE. So Dickson type design is suitable for applications where the system base on wake-up circuitry but has high requirement for the operating range, while the cross-coupled designs fit systems that rely on continuous energy harvesting.

For Dickson type architectures, [79] and [2] achieved -32 and -26.5 dBm sensitivity for OC condition. However their designs require zero-threshold transistors that are only available in special processes with additional cost, and have $< 10\%$ efficiency. [79] utilized off-chip air core high- Q inductors, together with 50 stages to enhance the output voltage, which increases the charging time to 155 ms. Therefore the design cannot be used for applications with high update rate.

For cross-coupled architectures, [5] demonstrated a 86% PCE with -19.2 dBm sensitivity by using feedback diodes for leakage reduction. However the PCE is obtained assuming perfect impedance matching throughout the dynamic range. To achieve the reported PCE, a variable matching network with high- Q values larger than 25 is needed.

Among all the architectures, the tunable- V_{th} RDR in this work achieves the lowest $V_{turn-on}$ of 140 mV with process variation cancellation and individual diode tuning. The tuning of V_{th} is performed in-house, separated from the rectifying operations. Combining the RDR with passive input voltage boosting in the matching network, a -27 dBm sensitivity and 22% PCE is predicted to generate 0.46 V output voltage for 500k Ω output load with a $Q = 10$ matching network. The improved performance of the proposed RDR comes at the cost of

area, which is noticeably larger than other rectifiers except for [55], which is yet to be optimized. The RDR in this work can be used for applications with strict sensitivity and power constraints but are flexible with the area.

5.5 Conclusion

A novel RF-to-DC RDR based on the tunable- V_{th} diode was presented, where each diode can be tuned at the wafer testing stage to the optimal values. An optimization algorithm was developed to condition the RDR automatically. The experimental measurement show $> 4dB$ improvement in sensitivity compared to the zero- V_{th} RDR with as small as 140 mV of $V_{in,peak}$ and a 500k Ω load. The proposed technique enables the adoption of more generic logic CMOS processes for RDR as well as provides an effective countermeasure to process variation.

Table 5.1: Summary Of Rectifier Performance and Comparison with Prior Arts

	Architecture	Freq (MHz)	Tech-nology	# of stages	Measured $V_{turn-on}$	R_L	V_{out}	Sensi-tivity (dBm)	PCE	Special require-ments	Compen-sation	Variation toler-ance	Area (mm^2)
Dickson	With zero- V_{th} , this work	541	0.18 μm	8	200 mV	∞	600 mV	-24	n.a.	Zero- V_{th} transistor, $Q = 10$	-	No	0.14
	With zero- V_{th} [79]	915	0.13 μm	50	-	∞	1 V	-32	n.a.	Zero- V_{th} transistor, air-core inductor with $Q \geq 50$	-	No	0.072
	With zero- V_{th} [2]	798	0.18 μm	6	150 mV	∞	1 V	-26.5	n.a.	Zero- V_{th} transistor, $Q = 10$	-	No	0.016
	V_{th} cancellation by auxiliary rectifiers [30]	433	0.18 μm	4	-	1M Ω	800 mV	-20	7.5%	-	during start-up	No	0.015
	Floating gate transistors [55]	910	0.25 μm	36	256 mV ^a	1.32M Ω	1 V	-18	15%	$Q = 6.4$	in house	No	0.4
	Tunable V_{th}, this work	570	0.18 μm	8	140 mV	500k Ω	460 mV	-27 ^b	22% ^b	$Q = 10$	in house	Yes	0.47
Cross-coupled	Cross-coupled with self bias [53]	953	0.18 μm	1	475 ^a	100k Ω	720 mV	-16.0	40%	$Q = 14$	during start-up	No	0.017
	Cross-coupled with custom antenna [103]	868	90 nm	5	-	1M Ω	1 V	-23	19%	non-50 Ω antenna	during start-up	No	0.029
	Dual-path with low- V_{th} transistor [64]	900	65 nm	5	300 mV ^a	147k Ω	1 V	-16.4	36.5%	Low- V_{th} transistor	during start-up	No	0.048
	With feed-back diodes for leakage reduction[5]	433	0.18 μm	1	-	100k Ω	1 V	-19.2	86%	$Q = 25$ to 35	during start-up	No	0.0084
Addi-tional tech	With in-ductive antenna[46]	2400	65 nm	5	-	1.8M Ω	1.6 V	-34.5	n.a.	$Q = 120$ inductive antenna and wake-up circuits	during start-up	No	0.5
	With boost converter [113]	900	0.18 μm	1	-	3M Ω	1 V	-26	60%	followed by booster converter	during start-up	No	-

^a $V_{turn-on}$ to achieve the specified V_{out} with given load R_L . In these references, passive input voltage boosting was used in the impedance matching network. The input voltage seen at the rectifier was calculated assuming perfect matching.

^b Based on the measured $V_{turn-on}$ of 140 mV by using a $Q = 10$ matching network in the simulation.

CHAPTER 6

EXPLORATION OF THE TUNABLE DIODES USED IN CROSS-COUPLED RECTIFIERS

In the previous chapter, we have compared the performance of the Dickson tunable- V_{th} rectifier with the prior works. From the comparison we can observe that generally Dickson architectures offer smaller $V_{in,peak}$ and better sensitivity, whereas cross-couple architectures demonstrate higher PCEs. Cross-coupled architectures usually utilize regular MOS transistors with about 0.5 V threshold voltages. The cross-coupled circuits can shift the diode IV curve dynamically, reducing V_{drop} in the positive cycle and suppressing the leakage current in the negative cycle, both leading to a higher PCE. However the turn-on voltage of the cross-coupled rectifier is still limited by V_{th} , meaning a higher than 0.5 V $V_{in,peak}$ is required at the input of the rectifier to power up the system, causing the worse sensitivity.

So is there a rectifier design that can achieve good PCE without hurting the sensitivity? To answer this question, we explore the utilization of tunable diodes in cross-coupled rectifiers, because they can improve the sensitivity without the need for special zero- V_{th} or low- V_{th} transistors.

6.1 Design of cross-coupled rectifiers using tunable diodes

In this section, we will discuss the design trade-offs for the cross-coupled rectifiers using tunable diodes. The conventional cross-coupled rectifiers using foundry provided transistors are also implemented for performance comparison. The schematics of a single stage of the fully cross-coupled design is shown

in Fig.6.1 [53, 5].

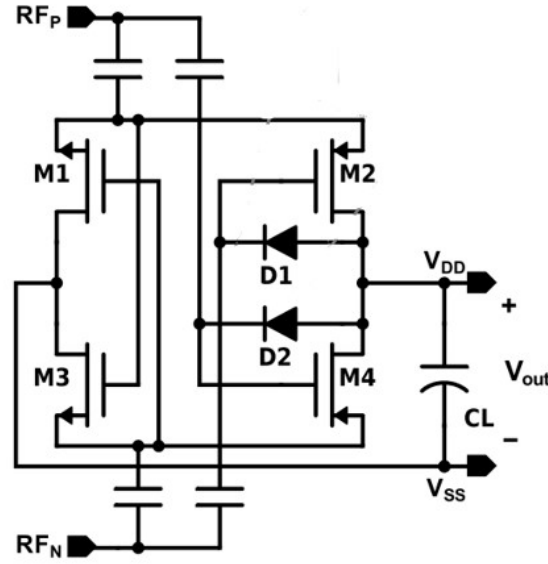


Figure 6.1: Schematic of a single stage for the cross-coupled rectifier. Cited from [5].

A 3-stage tunable cross-coupled rectifier was implemented, in which the transistors M1-4 in 6.1 are replaced by tunable diodes. Unlike Dickson type rectifiers where the PCE increases with stage numbers N for small N , and only starts to decrease when the N is bigger than 10. In cross-coupled designs, PCE decreases with increasing stage numbers, even if there are only 3 stages [53]. Therefore ideally we would want to use single-stage designs. However there is a trade-off between sensitivity, PCE and V_{out} in the selection of the stage number N . In tunable rectifiers, the improved sensitivity comes from lower $V_{in,peak}$ (given the same rectifier input impedance). But the biggest DC output voltage V_{out} is $2 * V_{in,peak}$ per stage (this is an over-estimate which doesn't consider leakage or output current to the load). When $V_{in,peak}$ is lower than 200 mV (correspond to -24 dBm sensitivity), a single stage cannot provide > 0.5 V output voltage. A stage number of 3 is chosen here as a compromise to provide sufficient V_{out} to

the load.

For the conventional cross-coupled rectifier, a single-stage one was implemented for better PCE. The component parameters used in the rectifier designs are summarized in Table.6.1. An output load $R_L = 100k\Omega$ was used for both designs.

Table 6.1: Parameters for the conventional and the tunable cross-coupled rectifiers

Parameter Name	Conventional	Tunable
C_c	550 fF	
C_L	1.13 pF	
Transistor L	180 nm	425 nm
NMOS W	$3.6 \mu m$	
PMOS W	$18 \mu m$	
C_d	-	10 pF
ΔV_{th} for NMOS	-	0.3 V
ΔV_{th} for PMOS	-	-0.4 V

In [5], the reported about 80% PCE only considered the rectifier assuming perfect impedance matching. Yet in reality there are more complications. First of all, the input impedance of a rectifier changes with input power level. It is challenging to achieve perfect matching over the entire input power range, unless a variable-Q matching network is implemented. The variable-Q matching network can be an array of different Q elements that are switched in and out in a programmable way at different input levels. But this adds to the design complexity and cost. Besides, the input resistance of the conventional rectifier varies from 33k to 64k Ω , which requires a Q value of 25 to 35, which is difficult to get on-chip. To take into consideration the above factors, a matching network

with a fixed Q of 15 was implemented in the rectifier and matching co-design. The total PCE, as well as matching efficiency and the rectifier PCE are studied for comparison.

6.2 Performance comparison

The simulated DC output voltages V_{out} as a function of $V_{in,peak}$ for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes are depicted in Fig.6.2. As can be seen, the tunable rectifier shows a lower $V_{in,peak}$ than the conventional one, because of the reduced V_{th} from compensation by P/E. To achieve $V_{out} = 400\text{mV}$, the lowest $V_{in,peak}$ for the tunable and the conventional rectifiers are 300 mV and 600 mV, respectively.

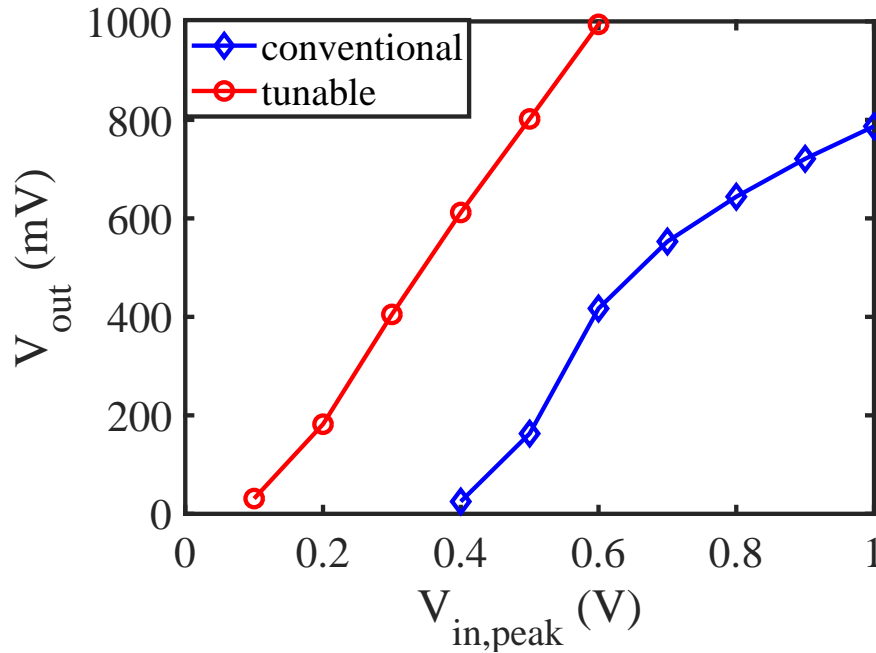


Figure 6.2: Simulated DC output voltage V_{out} as a function of $V_{in,peak}$ for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes.

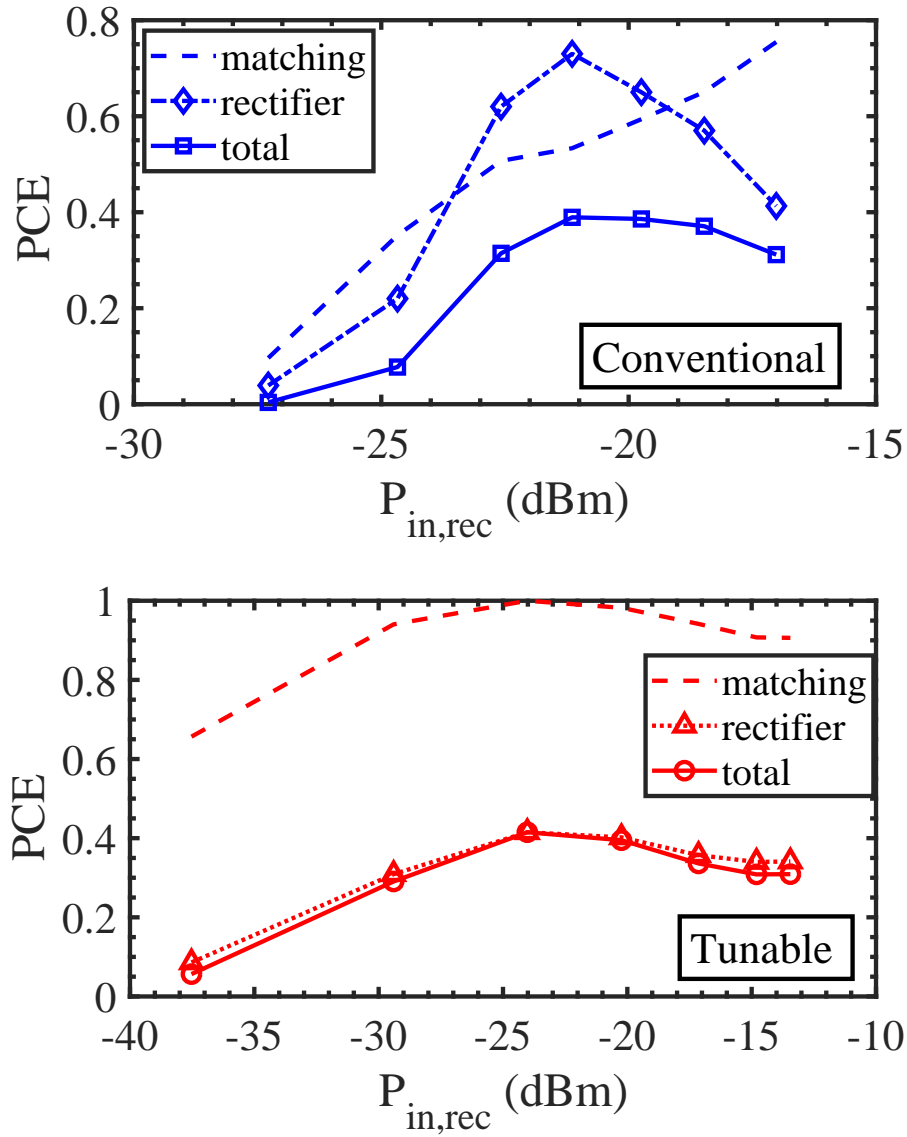


Figure 6.3: Total PCE and its components as a function of input power, for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes. The solid line with circular markers, the dashed line and the dash-dot line represent the total PCE of the system, the PCE in the matching network and that in the rectifier.

The total PCE and its components as a function of input power are plotted in Fig.6.3, for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes. Here, the solid line with circular markers, the dashed line and the dash-dot line represent the total PCE

of the system, the PCE in the matching network and that in the rectifier. For the tunable rectifier, the input resistance range is from 6k to 11.5 k Ω , which requires about 10 to 15 Q value. This Q value range is feasible for on-chip circuits which gives rise to a better efficiency in the matching network. For the conventional rectifier, peak total PCE of 40% is achieved at -21 dBm sensitivity. For the tunable rectifier, peak total PCE of 40% is achieved at -24 dBm sensitivity.

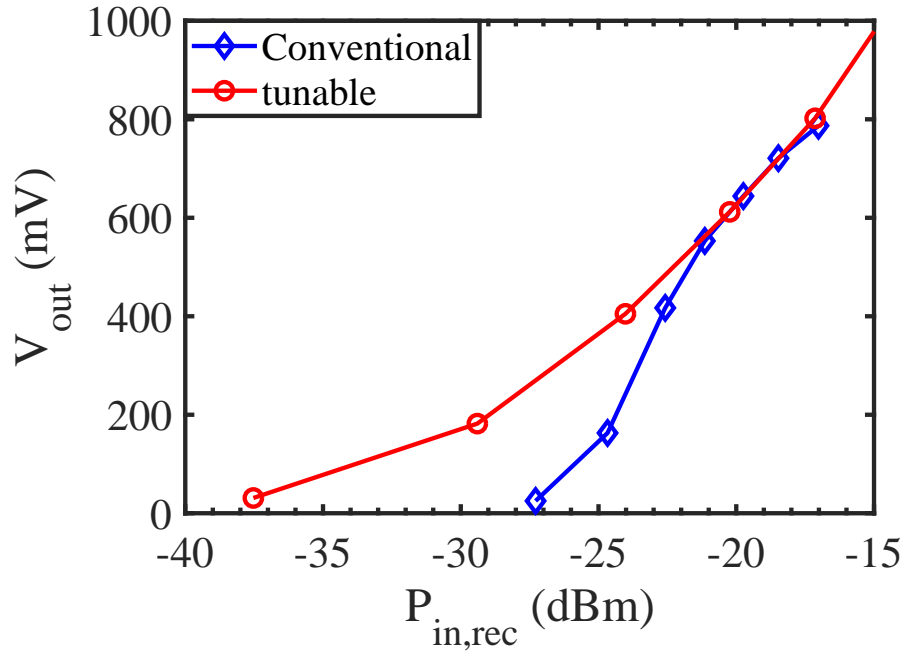


Figure 6.4: Simulated DC output voltage V_{out} as a function of the input power for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes.

Figure 6.4 demonstrates the simulated DC output voltage V_{out} as a function of the input power for the 1-stage conventional rectifier (blue line) and the 3-stage cross-coupled rectifier (red line) based on tunable diodes. Generally the tunable rectifier provides better sensitivity for lower V_{out} . The improvement in the sensitivity from the tunable structure becomes marginal when the input power level is high. At -24 dBm sensitivity, the tunable rectifier provides about 400 mV output for 4 μ A current. In contrast, the conventional one provides

about 200 mV output for 2μ A at the same input power level. The trend in the $V_{in,peak}$ versus V_{out} plot is different from that in the $P_{in,rec}$ versus V_{out} plot because the input power also depends on the input impedance, which varies with the power level.

6.3 Summary

The trade-offs in the design of cross-coupled rectifiers using tunable threshold diodes are summarized as below:

(1) For the same $V_{in,peak}$, we need to increase stage number for higher V_{out} . But PCE decreases with more stages.

(2) Big input impedance of the rectifier is needed for passive input voltage boosting in the matching. The input impedance decreases with stage numbers (because the stages are in parallel).

(3) To achieve the same input impedance, we need to increase transistor length, which costs more area and more charging time. And the input impedance can only be increased to a certain extent.

In the simulation in this dissertation, a 3 dB improvement in the sensitivity can be achieved by the tunable cross-coupled rectifier over the conventional one, with comparable PCE values. In order to design a RF-to-DC rectifier with even better sensitivity and PCE, a more complete study that includes the antenna, matching network and rectifier co-design is required. The design may need to go through a couple rounds of iterations and simulations to obtain the best performance.

CHAPTER 7

VARIABILITY IN ENTERIC NEURAL RECORDING: SOURCE AND RESOLUTIONS

7.1 Introduction to the Enteric nervous system and its variability

Enteric nervous system (ENS) is composed of 200-600 million neurons found in the gastrointestinal tract and plays a vital role in upholding gut functions of motility, epithelial secretion, and intestinal barrier [97, 28]. Disruption of ENS can cause Crohns disease, diabetes [118], irritable bowel disease (IBD) [105], and ulcerative colitis [8]. Different from those in the central nervous system (CNS), which is situated in the largely stationary brain, neural signals in the ENS are coupled to gut motility and peristalsis, interacting with gastrointestinal endocrine cells [107] and intestinal longitudinal muscles [111, 42]. These movements not only alter neural activities, but also cause electrode drift during recordings. Activities of the ENS is also influenced by the immune system via cytokines and mast cell tryptase [20, 62]. Furthermore, the ENS consists of various types of enteric neurons embedded in dense mesh-like 2-D ganglia, called myenteric and submucosal plexuses, and hence an extracellular electrode will pick up signals from an ensemble of heterogeneous neurons with different action potential waveforms. These factors render ENS recordings large waveform variability, which is compounded by the lack of knowledge of the ENS. To reliably interpreting the ENS extracellular action potential (EAP) recording, a spike classification method with high nonstationary tolerance is needed.

7.2 Prior works in spike classification for Enteric neural recording

The recent decade has seen enormous progress in spike recognition and classification methods for CNS. Such methods typically include three steps: (1) detection of candidate spikes, (2) waveform feature extraction, and (3) clustering of spikes. Different combinations of feature extraction, e.g. principle component analysis (PCA) [3], wavelet transform [90] and first and second derivative extrema (FSDE) [83], and clustering, e.g. k-means [16], superparamagnetic [90] and hierarchical adaptive means (HAM) [83], have been proposed for better accuracy. Overlapping spikes can be handled by either iteration and subtraction [59, 121, 87, 86, 24] or independent component analysis [109, 108, 44]. Most of these methods are off-line, requiring the collection of all spikes before running the analysis, unsuitable for real-time closed-loop applications. Additionally, these methods require off-chip processing due to the high computational cost, limited by the stringent power density on chip for safe neural implants [48].

A special case of the spike classification method for CNS that can be done in real-time is the template matching (TM) method, which uses the waveform in time domain directly in contrast to the extracted feature. TM compares each candidate spike to the template library and makes assignment by the largest similarity. The template library must be formulated in advance by feature extraction and clustering either from the initialization phase of the same experiment [75] or from earlier experiments. The key step in TM is the similarity calculation, for which Euclidean distance [94, 47] and cross correlation [48] are

most widely used. Both similarity calculations can be implemented in neural implant with real-time, low-power on-chip processing to render a reduced data transmission rate [63, 93, 115].

While many of these spike classification methods are used in CNS, they cannot be used as effectively in ENS. Minimal waveform variability in phase, duration and magnitude was often assumed, which is invalid due to electrode drift and non-stationary waveforms [7]. Electrode drift is caused by the neuron movement relative to the recording electrode, as well as the change in the electrolytic property of the biological environment [89]. Non-stationary waveforms refer to the change of the spike shape over time [116]. In CNS, Such variability problems have been investigated previously by modeling the source neurons as a mixture of Gaussians. Bayesian clustering [7] was employed to calculate the candidate clusters in short-time frames, and the transition probabilities among each cluster mixture determine the cluster choice as the maximum-a-posteriori solution. Simplification based on Kalman filtering [13] can be utilized to enhance efficiency. Bayesian optimal TM [25] was also suggested for online applications. However, the main focus of these methods was on the variability of the spike magnitude caused by the movement of the cluster centers, but not on the time course, which proves critical for ENS EAP recordings due to large colony motility [111] in the complex environment.

7.3 Proposed new classification method for neural recording with large variability

Here we propose dynamic time warping (DTW) as a similarity measure with high tolerance to the spike variability in both magnitude and time. DTW has proven to be very effective in speech recognition under source and ambient variations [122] and ECG profile characterization [40], which are similar to the ENS EAP analysis at hand. We established a fastDTW method with automatic thresholding for ENS recordings with large spike variability for real-time applications. During the similarity calculation, we have also implemented adaptive temporal gridding [96] with linear complexity. The performance of fastDTW is first evaluated on synthesized ENS EAP data at various noise levels, showing remarkable enhancement in accuracy and computational complexity in comparison to cross-correlation based TM (CCTM) and PCA + k-means clustering without time warping. To demonstrate the manner and degree how our fastDTW method can impact neural recording, experimental EAP of the mouse enteric neurons under different stimuli are analyzed, where our fastDTW method was able to successfully classify biphasic and monophasic spikes when the waveform variability is more than millisecond in width and millivolt in magnitude. In addition to more precise spike counting than the TM method, fastDTW also directly provides specific waveform parameters with associated changes in time for further statistical processing and behavior recognition.

CHAPTER 8

FAST DYNAMIC TIME WARPING SPIKE CLASSIFICATION

ALGORITHM: METHOD AND PERFORMANCE

8.1 introduction

This chapter describes the proposed fastDTW algorithm for spike classification in high noise and high variability environment. Sections 8.2.1 and 8.2.2 introduce the theory and procedures of fastDTW algorithm. Section (8.3.1) establishes the accuracy benchmark for spikes classification algorithms. The performance of fastDTW is then analyzed in details in terms of classification accuracy and computational complexity in Section (8.3).

8.2 fastDTW algorithm

The procedures to perform fastDTW classification is described in this section. We will first introduce dynamic time warping (DTW) as a similarity measure with high tolerance to the spike variability. Automatic thresholding and fast-DTW are then added to enhance real-time performance and to complete the algorithm.

8.2.1 DTW in similarity calculation

We will show how DTW can be effectively and efficiently employed for similarity calculation in the neural spike classification algorithm when large variability

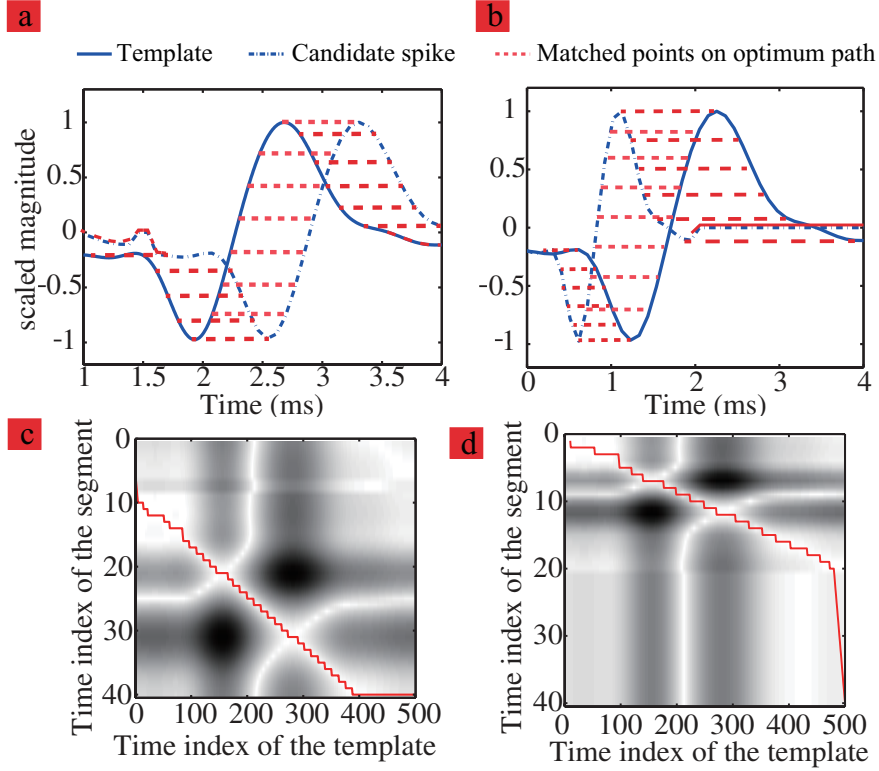


Figure 8.1: Similarity calculation using direct DTW. (a)(b): Template (blue solid lines) and candidate spike (blue dash dot lines) with the matched points connected by red dashed lines. (c)(d): Similarity matrices with the minimum-distance warping paths (red) for the spikes in (a) and (b).

is expected.

Let X denote the time series of the spike template with length n , and Y denote the time series of the candidate spike segment with length m (X and Y are normalized in magnitude before similarity calculation):

$$X = x_1, x_2, \dots, x_i, \dots, x_n \quad (8.1)$$

$$Y = y_1, y_2, \dots, y_j, \dots, y_m \quad (8.2)$$

where i, j are time indexes of the series X and Y , respectively.

The similarity matrix $[a]_{nm}$ is constructed where each element is the normal-

ized Euclidean distance between a pair of points in series X and Y :

$$a_{ij} = d(x_i, y_j) = \frac{|x_i - y_j|}{m + n} \quad (8.3)$$

A warp path P is defined on the similarity matrix $[a]_{nm}$

$$P = p_1, p_2, \dots, p_k, \dots, p_l \quad (8.4)$$

where l is the length of the warp path and the k^{th} element of the warp path is

$$p_k = (i, j) \quad (8.5)$$

The warp path starts at $p_1 = (1, 1)$, and ends at $p_l = (n, m)$, with the time indexes increasing monotonically:

$$p_k = (i, j), \quad p_{k+1} = (i', j') \quad i \leq i' \leq i + 1 \quad j \leq j' \leq j + 1 \quad (8.6)$$

The minimum-distance warp path P is then found, with a cumulative distance of:

$$d(P) = a_{11} + \sum_{k=1}^{K-1} d(p_k, p_{k+1}) = \sum_{p_k=(i,j) \in P} a_{ij} \quad (8.7)$$

The minimum-distance warp path P in Eq.8.7 is usually obtained by dynamic programming:

$$D(i, j) = a_{ij} + \min(D(i, j - 1), D(i - 1, j - 1), D(i - 1, j)) \quad (8.8)$$

The minimum $d(P)$ is between 0 and 1, which is used as an indicator of the similarity ξ between the template and the candidate spike :

$$\xi = 1 - d(P) \quad (8.9)$$

Based on the minimum-distance warp path P , a new time series Y' with length n can be created, which is the candidate spike time warped (TW) to the template:

$$Y' = y_{s_1}, y_{s_2}, \dots, y_{s_h}, \dots, y_{s_n} \quad (8.10)$$

The time index s_h of the time warped candidate spike Y' is the 2nd coordinate of the points on P , when the 1st coordinates go from 1 to n :

$$S = s_1, s_2, \dots, s_h, \dots, s_n, \quad (h, s_h) \in P, \quad h = 1, 2, \dots, n \quad (8.11)$$

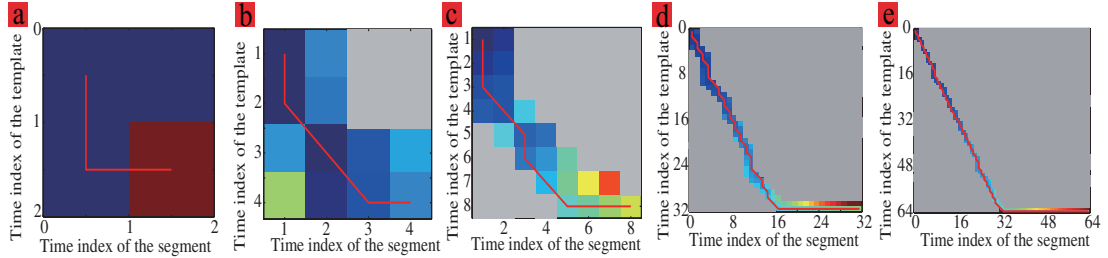


Figure 8.2: FastDTW expedites direct DTW to linear time complexity. At each resolution stage, the optimal warping path is found in a window (colored square) containing the lower resolution warping path (red solid line).

Figure 8.1 shows an example of the similarity calculated with DTW. The candidate in Fig. 8.1(a) is the template parallel-shifted by 25% of the segment width. If we denote the segment width to be 360° in phase, a 25% shift corresponds to a 90° phase shift. The candidate spike in Fig. 8.1(b) is the template scaled to half the duration. The similarity matrices are plotted in Figs. 8.1(c)(d), respectively. Darker color indicates a larger Euclidean distance and less similarity. The minimum-distance warp path P is plotted in a red line on top of the similarity matrix. P consists of two parts: the distance from the waveform similarity independent of the warping in time domain, and the distance from spike misalignment, stretching or shrinking. The similarities ξ in Figs. 8.1(a)(b) are 0.945

and 0.95, respectively, very close to 1 for exact matching. The matched pairs of points in the template and the candidate spike are connected by red dash lines.

DTW in Eq. 8.8 has a complexity of $O(N^2)$, where N is the size of the input series. In order to expedite the algorithm, we adopted fastDTW in our work ([96]) with a linear complexity $O(N)$. FastDTW accelerates DTW by coarsening and refinement, as demonstrated in Fig. 8.2. P is first found in the lowest resolution by coarsening the two input series. The path searching returns to high resolution by refining the area containing the present path in a quad-tree manner. The search continues until the original series resolution is reached.

8.2.2 Overview of the fastDTW method with automatic thresholding

Now that we have introduced DTW as the similarity measure with high tolerance to variabilities, we continue to present the fastDTW algorithm for real-time neural spike classification. The fastDTW algorithm has three main parts as shown in Fig. 8.3: pre-processing for one time before each recording setup, spike candidate detection, and spike classification by fastDTW as a similarity measure. The latter two steps can be pipelined to decrease the overall computational time.

Pre-processing

First, the recorded data are passed through a bandpass filter from 1 Hz to 30 kHz in Fig. 8.3(a) to remove the slow voltage drift and high frequency noise.

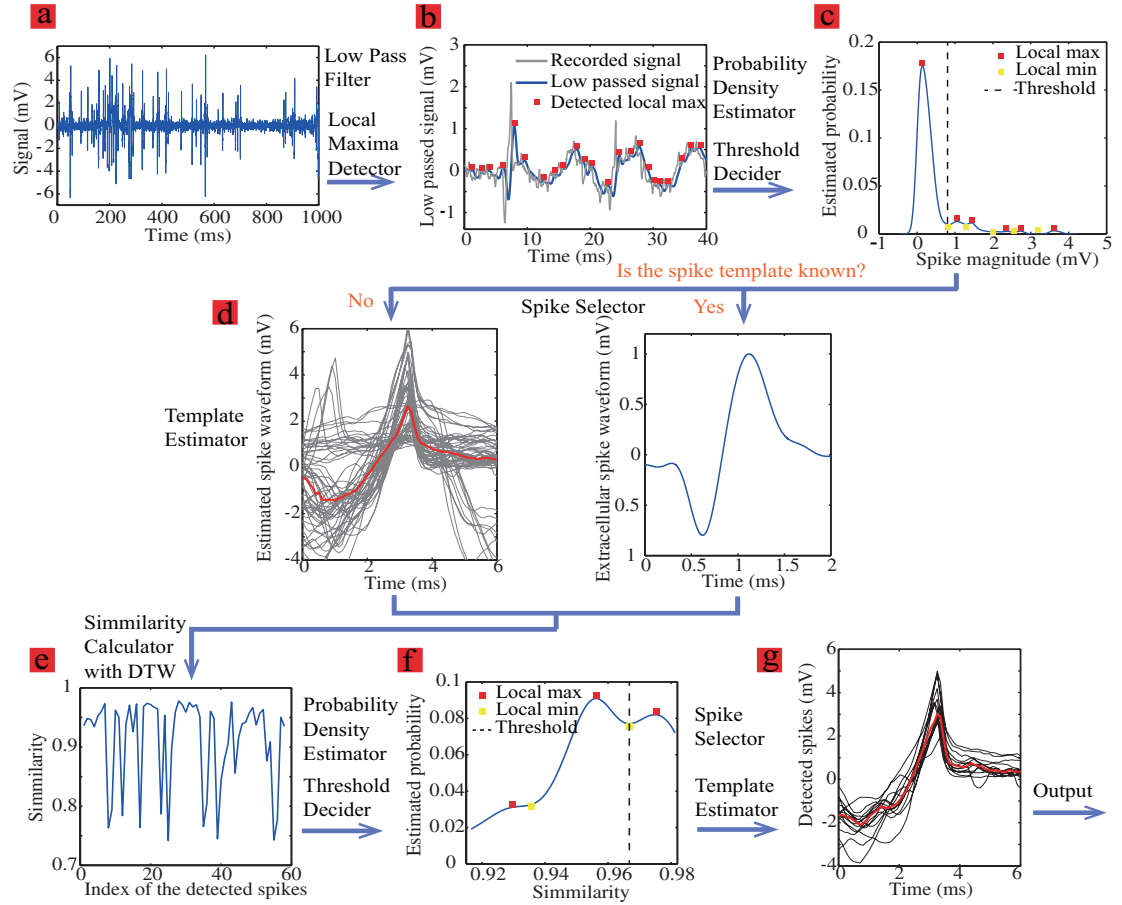


Figure 8.3: Flow chart of fastDTW with automatic threshold decision. (a) Band-passed data of EAP. (b) Low-passed (100Hz) waveforms (blue) on top of the raw signal (grey) with the local maxima (red) detected as candidate spikes. (c) Magnitude distribution of candidate spikes, with a threshold determined as the local minimum to the right of the largest maximum. (d) Spike template used in classification. With no available templates (middle left), the estimated spike templates (red) are obtained by averaging the candidate spikes (grey) above the threshold in the first few time bins during initialization. (e) The similarities between the candidate spikes and the template are calculated by fastDTW. (f) The automatic similarity threshold is determined in the same way as in (c). Candidate spikes with larger similarity than the threshold are assigned to the template group. (g) A new averaged spike waveform is calculated from the classified spikes in the time bin, which can modify the previous spike template adaptively to account for the continuously slow change.

Then the filtered data is cut into short time bins to collect statistics for further automated processing. The bin size needs to be sufficiently large to expect to encompass more than 10 spikes under stimulation for better statistical collection

and automated decision making. Although the choice of bin location and size can readily be adaptive, we have used a fixed size of 250 ms for the synthesized data, and 500 ms for the experimental data for simplicity in this work. Different bin size can also be chosen depending on the application. The spikes in each time bin will then be detected and classified.

Spike candidate detection

The candidate spikes in each time bin are detected by automatic thresholding. The threshold is determined by the statistics of local maxima. The magnitude distribution of the local maxima is extracted for its probability density in Fig. 8.3(c). The largest maximum of the estimated probability is most likely due to noise, while the other maxima by actual EAP ([48]). The threshold is determined automatically by locating the first minimum to the right of the first maximum.

Candidate spikes with peak magnitude larger than the threshold are saved in segments, with the segment width that is larger than but close to the anticipated spike duration. If the spike duration cannot be easily estimated beforehand, or the duration varies significantly over time, multiple segments with different widths covering the duration range are saved for each spike candidate.

Although the peak location of the candidate spike is estimated in the threshold determination procedure, the candidate spike can still have temporal misalignment with the template. In conventional TM where cross-correlation or sum of Euclidean distance is used as the similarity measure, either peak alignment or a sliding window scheme needs to be applied according to the mis-

alignment tolerance. Accurate peak alignment to the template usually requires up-sampling ([94]), and is difficult in many situations. For example, if the spikes are aligned to the peak with the largest absolute value, biphasic spikes that have similar magnitudes for the positive and negative peaks can be easily misaligned. "Sliding window" saves multiple shifted versions of each candidate spike with equal width. Spike recognition is then performed on the multiple segments for each candidate. In our fastDTW method, which can tolerate large duration and alignment variation, omission of peak alignment or "sliding window" greatly reduces the computational complexity.

Spike classification

When no spike template is available, the template estimator first calculates the averaged spike waveform in the first few time bins from the saved spike candidates by taking the median of the spikes (Fig. 8.3(d)). The estimated templates are used to start the algorithm and are modified adaptively to deal with the evolution of spike shapes over time.

When the spike templates are known from cellular model construction ([26]) or feature extraction and clustering in earlier experiments, the candidate spikes in the time bins can be classified by similarity directly using fastDTW (Fig. 8.3(e)). The similarity distribution is formed with the probability density estimator and the similarity threshold (Fig. 8.3(f)) is automatically determined in the same way as the spike magnitude threshold. Spike candidates with similarities above the similarity threshold to a template are classified to that group, and the rest are further checked by other templates or are discarded. With the classified spikes in the time bin, the template estimator can iteratively modify

the template by learning over time (Fig. 8.3(g)). The classification process is repeated for all templates and then proceeds to the next time bin. If a spike candidate is recognized by more than one template group, it should be classified to the group with highest similarity.

A hard threshold of similarity can be alternatively set for each template to reduce the false positive rate as well. If the automatic similarity threshold by statistics is smaller than the hard threshold, the similarity threshold is forced to the hard threshold. In conventional TM method, the choice of the hard threshold greatly influence the accuracy. In contrast, fastDTW has high tolerance to the threshold variation and in most case the automatic similarity threshold is sufficient without the need for a hard threshold. Our proposed fastDTW algorithm is thus fully automatic, suitable for experimental recordings with pixel-array recording sites ([23, 44, 43]), where manual intervention of the threshold for each pixel would be impractical.

8.3 Performance analysis

To evaluate the effectiveness of the proposed fastDTW algorithm, its accuracy and complexity were compared with conventional template-matching method based on cross-correlation (CCTM) and PCA + k-means clustering methods with and without time warping. The results of CCTM and PCA + k-means clustering methods were implemented and reproduced for the comparison.

CCTM follows the same procedure in Section 8.2.2 except that cross-correlation in Eq.8.12 is used as the similarity measure instead of DTW. For each candidate spike, CCTM runs on 3 segment widths. At each segment width, slid-

ing window of 10 shifted segments is employed.

$$\xi = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (8.12)$$

For PCA + k-means clustering, all candidate spikes in the whole recording are collected, aligned to the largest peak and segmented before processing. Spike detection is done in a similar manner as described in Section 8.2.2. Candidate spike segments with or without TW are used as inputs of PCA. With TW, each candidate spike is time warped to the the most similar template using DTW before running PCA. The minimum distance warping path P between the candidate spike and the template is found as described in Section 8.2.1, and the candidate spike is time warped using Eqs. 8.10 and 8.11. The principle components are then utilized for k-means clustering.

The accuracy and complexity analysis of different methods are presented in Section (8.3.1) and (8.3.2).

8.3.1 Accuracy analysis

Accuracy is one of the most important metrics of a classification algorithm. Section (8.3.1) introduces the synthesized ENS recordings for accuracy benchmark with known ground truth. Section (8.3.1) compares fastDTW and CCTM in their abilities to deal with spike mis-alignment and spike duration variations in similarity calculation. Section (8.3.1) shows the accuracy comparison of different methods on the synthesized recordings.

Accuracy benchmark on synthesized EAP recording

Synthesized ENS EAP recordings are used for accuracy benchmark of different algorithms. The recording is synthesized as the sum of known EAP waveforms and white noises. Two spike waveforms are used in the simulation: biphasic and monophasic (Fig.8.4), which had been observed in ENS EAP experiments ([102, 18, 78]). The duration of the synthesized spikes follows Gaussian distribution with 1 ms mean and different variances are applied to mimic the duration variation in ENS. Each recording contains 100 biphasic spikes and 100 monophasic spikes at random firing time (spacing larger than the refractory period of 3 ms) at signal-to-noise ratios (SNR) from 10 to 3.

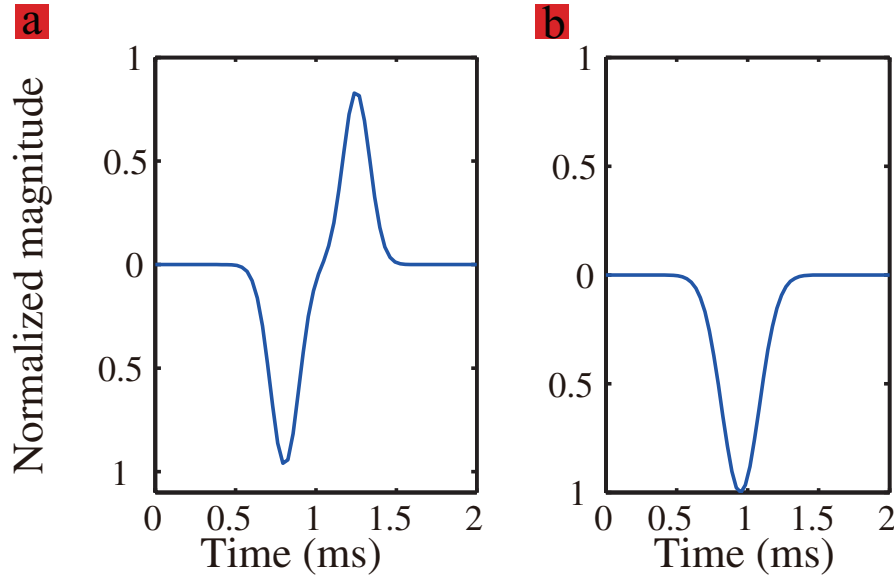


Figure 8.4: EAP waveforms of (a) the biphasic and (b) the monophasic spikes

The accuracy of the methods are evaluated by the true positive rates (R_{TP}) and the sum of false positive (FP) and false negative (FN) at different noise levels. When the method reports a spike of certain template group and the spike truly belongs to that group, a TP is recorded. Otherwise the reported spike is a FP. When a spike in the synthesized data is not reported by the method, a FN is

recorded.

$$R_{TP} = \frac{\text{number of TP spikes}}{\text{total number of spikes in the recording}} \quad (8.13)$$

R_{TP} is between 0 and 1, when 1 represents all spikes being correctly classified by the method. There are two sources for FP spikes: by the background noise mistaken as a spike, and by misclassification to the wrong template group. An accurate method provides a high R_{TP} and a small sum of FP and FN (FP+FN).

Spike alignment and duration variation on similarity

The effect of timing variability on similarity calculation and the resulting spike recognition ability is compared between fastDTW and CCTM. Two spike templates of biphasic and monophasic waveforms are adopted in the synthesized recording (Fig. 8.4). The candidate spike segments are of the same waveforms, either shifted to mimic the misalignment, or shrunk or expanded to mimic the duration variation (similar to Fig. 8.1).

In order to achieve correct spike recognition, the similarity between the candidate spike and the true template group must be larger than the similarity threshold, and the similarity with all other groups should be smaller than the threshold. In the ideal case without noise, the minimum similarity threshold for correct spike discrimination is 0.95 for fastDTW and 0.82 for CCTM (Fig. 8.5).

The similarity is a function of the spike misalignment, as calculated by fastDTW (Fig. 8.5(a) top) and by CCTM (Fig. 8.5(b) top). Using the similarity threshold criteria, fastDTW can correctly detect and discriminate both waveforms in a 201° window out of 360° for misalignment. In contrast, CCTM only allows a 18° window. In other words, at most 2 shifted segments are needed

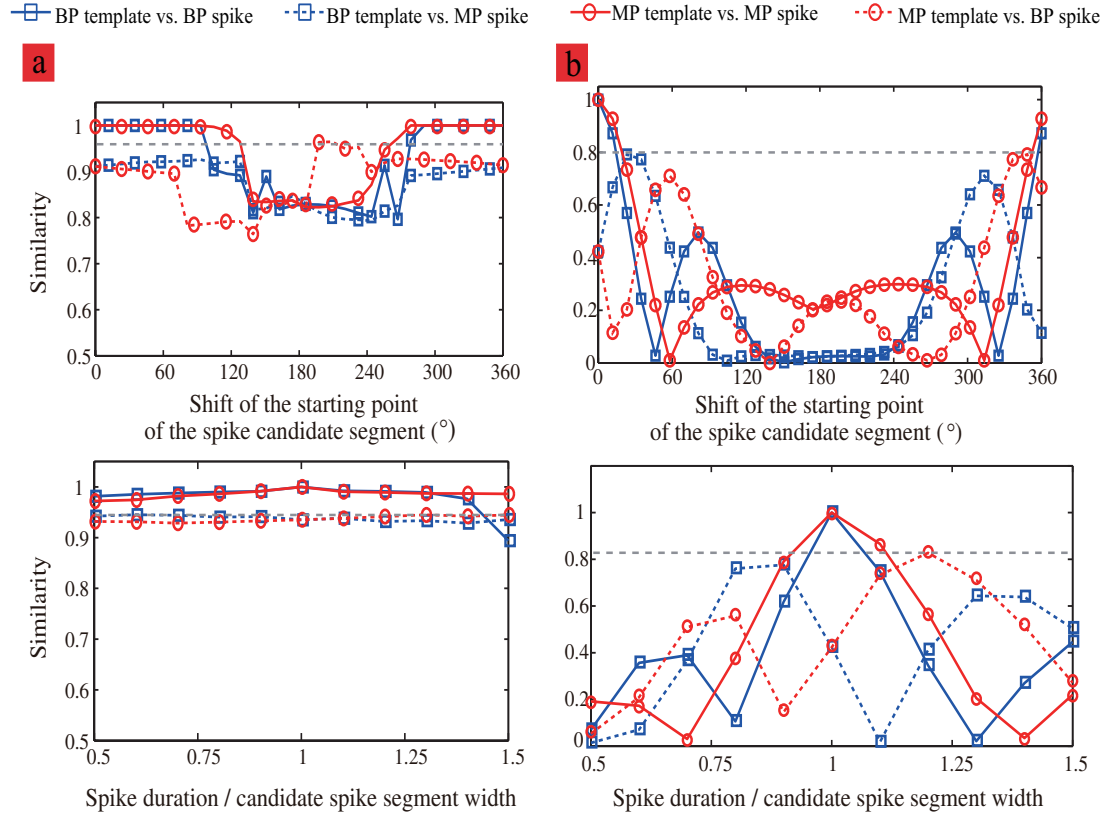


Figure 8.5: Comparison of fastDTW and CCTM in similarity calculation: (a) Similarity calculated by fastDTW with spike misalignment (top) and duration variation (bottom). (b) Similarity calculated by CCTM with spike misalignment (top) and duration variation (bottom). BP stands for biphasic, and MP stands for monophasic. The grey lines represent the minimal similarity threshold that ensures no misclassification in the ideal case without noise.

(and in many situations one segment works reasonably well) for each candidate spike by fastDTW, in comparison to 20 segments by CCTM. Because of the tolerance to misalignment, fastDTW is expected to have less computational cost than the CCTM to achieve the same classification accuracy, by avoiding repetitions on shifted segments.

The similarity is also a function of the spike duration variation, as calculated by fastDTW (Fig. 8.5(a) bottom) and by CCTM (Fig. 8.5(b) bottom). FastDTW can detect and discriminate the biphasic and monophasic spikes accurately in

mosy duration variation from 0.5 ms to about 1.45 ms. In comparison, CCTM can only tolerate the duration variation of about 0.2 ms for monophasic spikes and 0.12 ms for biphasic spikes. As a result, CCTM will need to run on more than 6 segment widths for the duration variation between 0.5 and 1.5 ms so as to achieve same accuracy as fastDTW.

algorithm accuracy

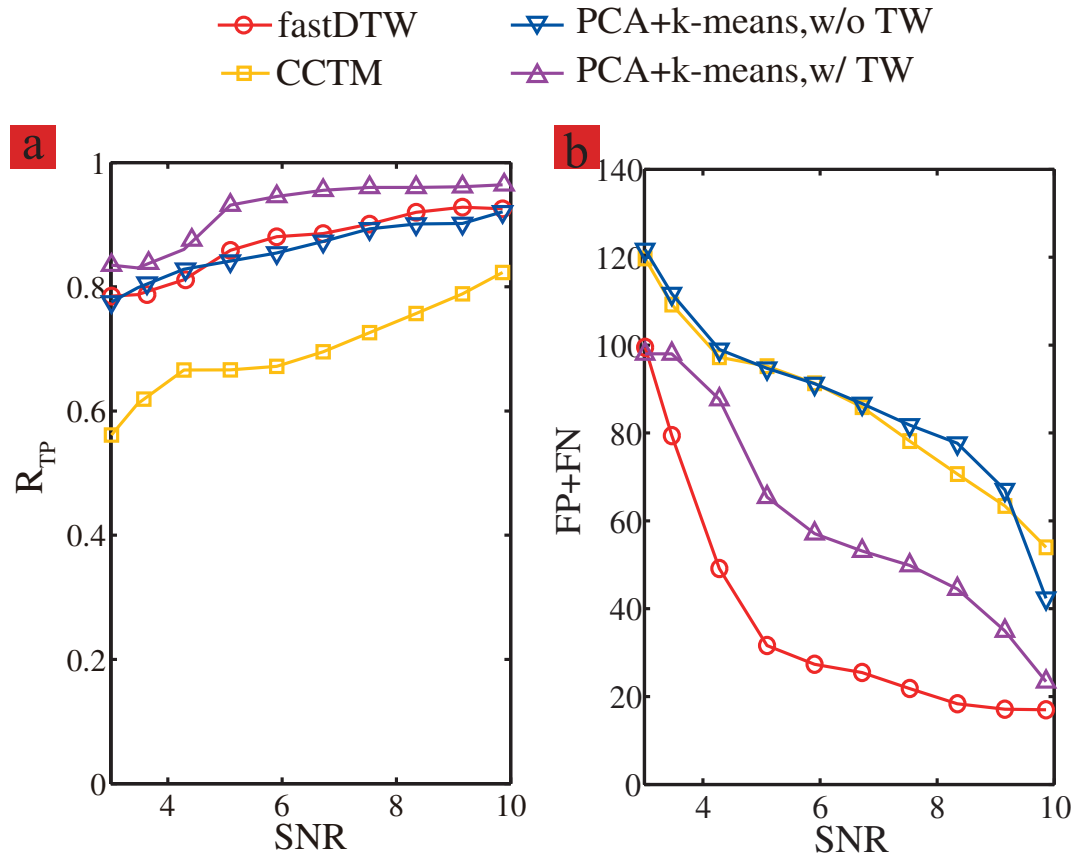


Figure 8.6: R_{TP} (a) and FP + FN (b) as a function of SNR by different methods. For fastDTW and CCTM, the similarity threshold is chosen to minimize the FP + FN. Spike duration $\propto \text{norm}(1,0.3)$ (Gaussian distribution) in the synthesized recording.

The accuracy of fastDTW is compared with those of CCTM and PCA + k-means clustering with and without TW as described in Section 8.3.1. The accu-

racy as a function of SNR is plotted in Fig. 8.6. FastDTW has a R_{TP} only slightly smaller than PCA + k-means clustering with TW, which is an off-line method requiring data collection of all spikes before processing. Among these methods, fastDTW shows the smallest FP + FN at all noise levels.

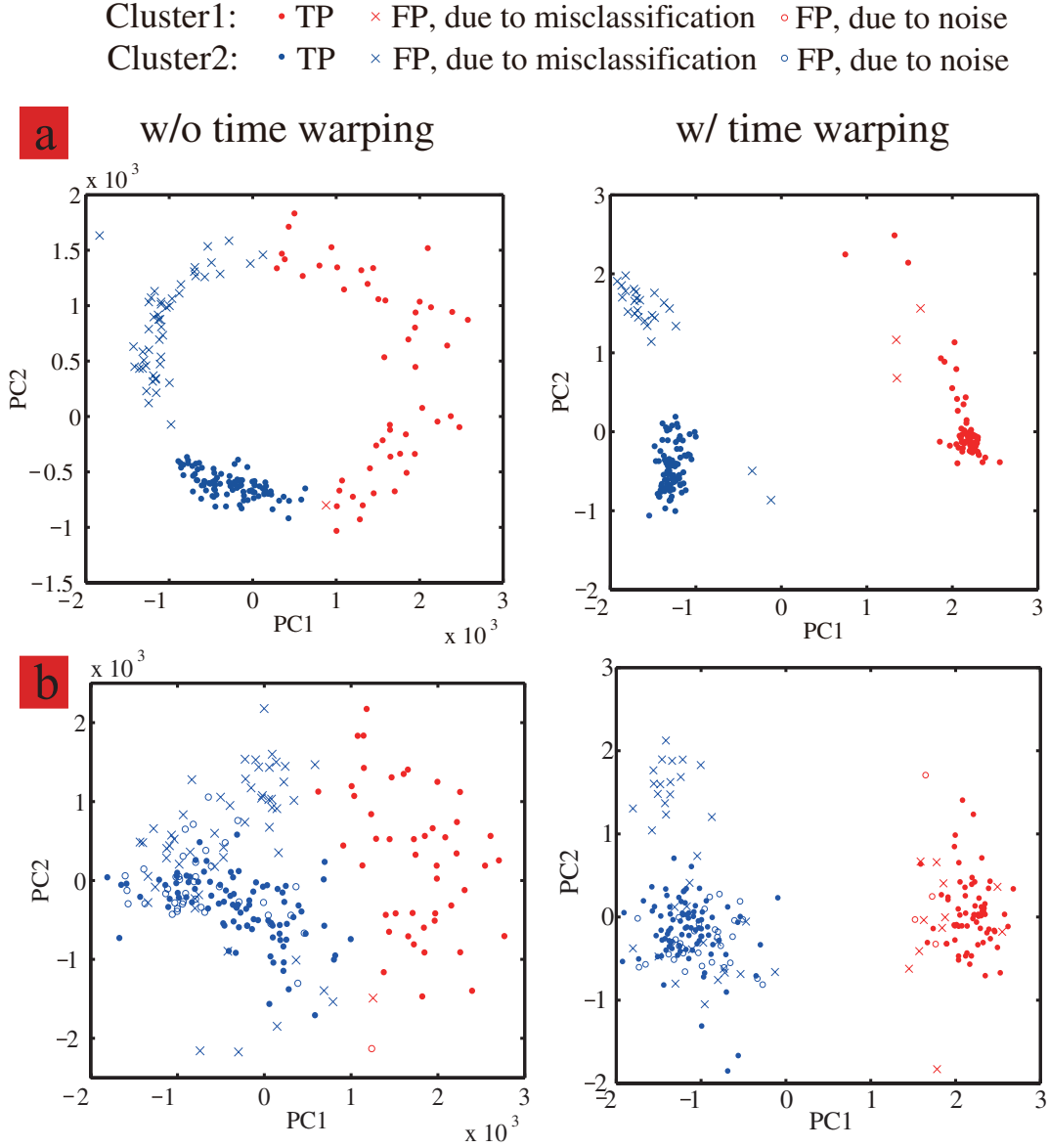


Figure 8.7: Classified spikes projected to the plane of the first two principle components, by PCA + k-means clustering with and without TW, at SNR=10 (a) and SNR=3 (b). Spike duration $\propto \text{norm}(1,0.6)$ in the synthesized recording.

Fig. 8.7 demonstrates how time warping of candidate spikes increases the

accuracy of PCA + k-means method at two noise levels SNR=10, and 3. The candidate spikes assigned to Cluster 1 (monophasic) and 2 (biphasic) are plotted in blue and red, respectively. Without TW, it's difficult to separate the two clusters even at SNR=10 because the shapes are non-Gaussian due to spike waveform variability (Fig. 8.7(a) left). The estimation of cluster centroids is rather hard and about 50% of biphasic spikes are misclassified as monophasic. The average waveforms of the true and misclassified spikes can be found in Supplementary Material. With TW, both R_{TP} and the FP + FN improved noticeably with clearly separable clusters (Fig.8.7(a) right). Most of the misclassified spikes are outliers far away from the cluster centroids, which may be removed by applying constraints on the largest distance to the centroids to reduce FP. At SNR=3, the clusters are further corrupted by noise (Fig. 8.7(b) left). However TW is still able to improve the accuracy of PCA + k-means clustering.

The waveform dependency of accuracy by different methods is depicted in Fig. 8.8. Using highly waveform dependent method, the classification of one (or a few) cluster(s) can have many errors, even the overall FP+FN is low. The variation of accuracy with different template waveforms by fastDTW is the smallest. With fastDTW, the biphasic spikes demonstrate slightly higher R_{TP} than the monophasic spikes ([76]).

The influence of the similarity hard threshold on accuracy is investigated for fastDTW and CCTM (Fig. 8.9). As described in Section 8.2.2, besides the automatically determined similarity threshold by statistics Th_{auto} , a hard threshold Th_{hard} can be alternatively set to reduce the FP rate. If Th_{auto} is smaller than Th_{hard} , the similarity threshold Th_{sim} is forced to the Th_{hard} in Eq.(8.14).

$$Th_{sim} = \max(Th_{auto}, Th_{hard}) \quad (8.14)$$

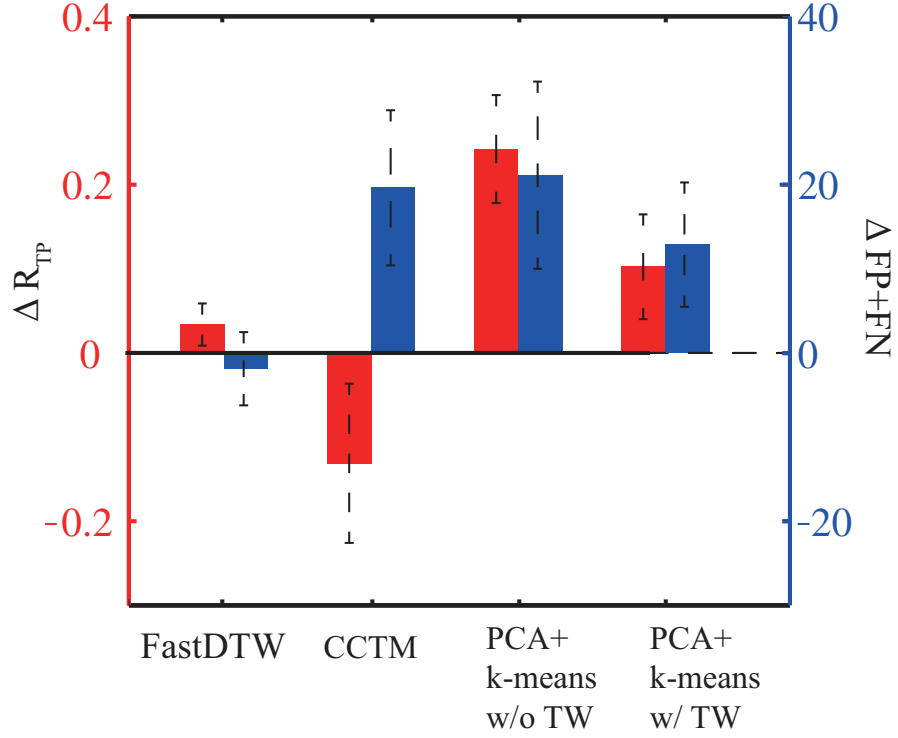


Figure 8.8: ΔR_{TP} (red) and $\Delta FP + FN$ (blue), between biphasic and monophasic spikes, by different methods. Spike duration $\propto \text{norm}(1,0.3)$ in the synthesized recording.

The relative hard threshold is defined in Eq. (8.15).

$$\text{relative hard threshold} = \frac{Th_{\text{hard}} - Th_{\text{hard,min}}}{Th_{\text{hard,max}} - Th_{\text{hard,min}}} \quad (8.15)$$

Where $Th_{\text{hard,max}}$ and $Th_{\text{hard,min}}$ are as follows.

$$Th_{\text{hard,max}} = \min[Th_{\text{hard}} : TP \rightarrow 0] \quad (8.16)$$

$$Th_{\text{hard,min}} = \max[Th_{\text{hard}} : FP \rightarrow FP_{\text{max}}] \quad (8.17)$$

When Th_{hard} is as high as $Th_{\text{hard,max}}$, no spike is classified to the template group and the relative hard threshold is 1. When Th_{hard} is $Th_{\text{hard,min}}$, FP is

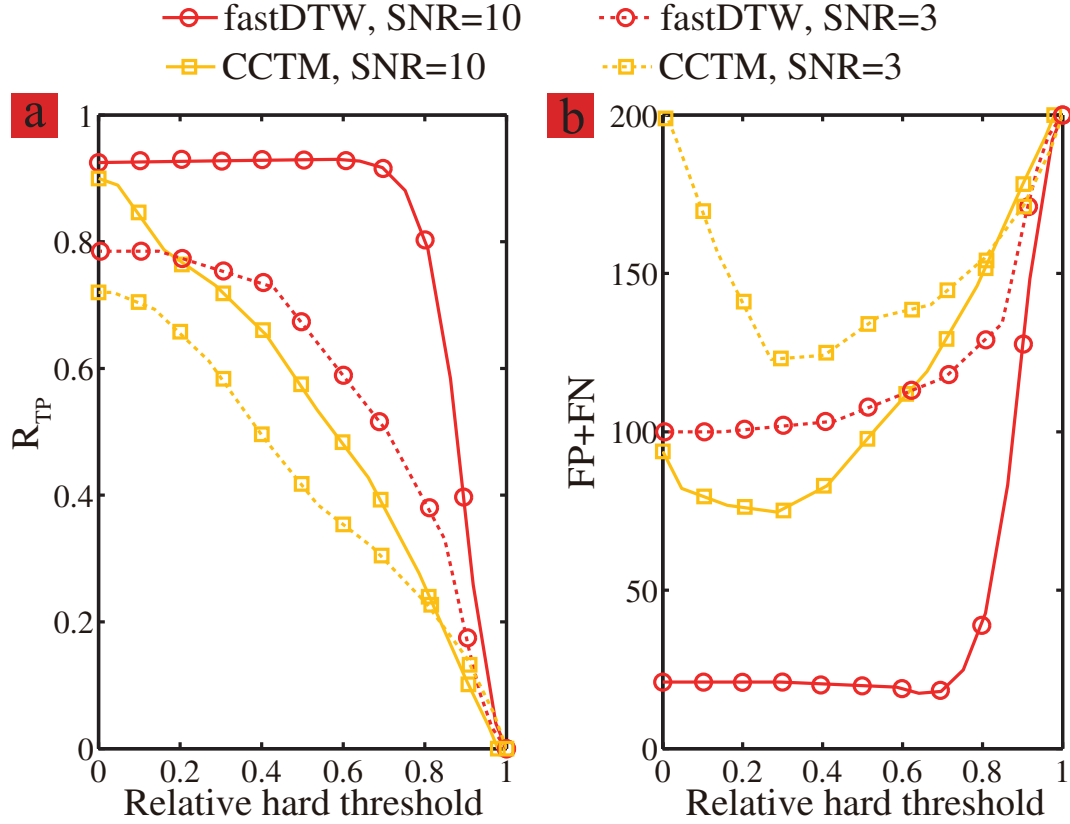


Figure 8.9: R_{TP} (a) and FP + FN (b) as a function of relative similarity hard threshold by fastDTW and CCTM. Spike duration $\propto \text{norm}(1,0.3)$ in the synthesized recording.

the largest and the relative hard threshold is 0. $Th_{\text{hard,max}}$ and $Th_{\text{hard,min}}$ are determined by sweeping Th_{hard} .

R_{TP} (FP + FN) decreases (increases) more slowly with increasing relative hard threshold by fastDTW than by CCTM. The accuracy is less sensitive to the change of the relative hard threshold by fastDTW. Even if it goes to zero, fastDTW still gives a R_{TP} of about 0.92 (0.8) and a FP + FN of about 20 (100) at SNR=10 (SNR=3). In other words, when fastDTW is used, TR_{auto} is sufficient to reject most false spikes, with no need to choose an optimized hard threshold. FastDTW is thus fully automatic, suitable for experimental recordings with pixel-array recording sites ([23, 44, 43]), where setting Th_{hard} manually for

each pixel would be impractical.

8.3.2 Computational complexity

The computational complexity is important to realize on-chip spike classification due to the stringent constraints on power, bandwidth and area. The computational complexity of fastDTW is compared with those of CCTM and PCA + k-means clustering in Table 8.1. The computational complexity is defined as the number of operations performed by the algorithm (eg. addition, multiplication) ([101]), where each operation takes a fixed amount of time to perform. PCA can be transformed to a problem of singular-value decomposition (SVD) of a matrix ([39]). The complexity of k-means clustering is estimated assuming Lloyds algorithm ([61]) which runs iteratively for the local optimum. For fastDTW and CCTM, the complexity is estimated on the similarity calculation, because other steps of the algorithm run only once for the whole time bin.

The actual running time is subject to the hardware platform and different schemes have been introduced to determine a single complexity figure-of-merit (CFOM). For example, [31] and [82] define CFOM as the sum of the number of additions and ten times the number of multiplications. Here we assume ARM Cortex M4 as the hardware platform and define CFOM as the weighted sum of operation numbers, where the weight is the number of clock cycles for each operation:

$$\text{CFOM} = N_{\text{addition}} + N_{\text{multiplication}} + N_{\text{shift}} + N_{\text{condition}} + 10N_{\text{division}} + 14N_{\text{squareRoot}} \quad (8.18)$$

Figure 8.10 demonstrates the trade-off between the error rates and CFOM. The value of the variables are shown in Table 8.2. It can be clearly seen fastDTW offers the lowest error rate with a very small complexity (3000 clock cycles per spike). Assuming a micro-controller frequency of 10 MHz, each spike is classified within 0.3 ms, making fastDTW a real-time method. Assuming 100 recording channels, each with a firing rate of 20 Hz, this corresponds to 6 MIPS and $2 \mu\text{W}$ /recording channel in power consumption (ARM Cortex M4 consumes $33 \mu\text{W}/\text{MHz}$).

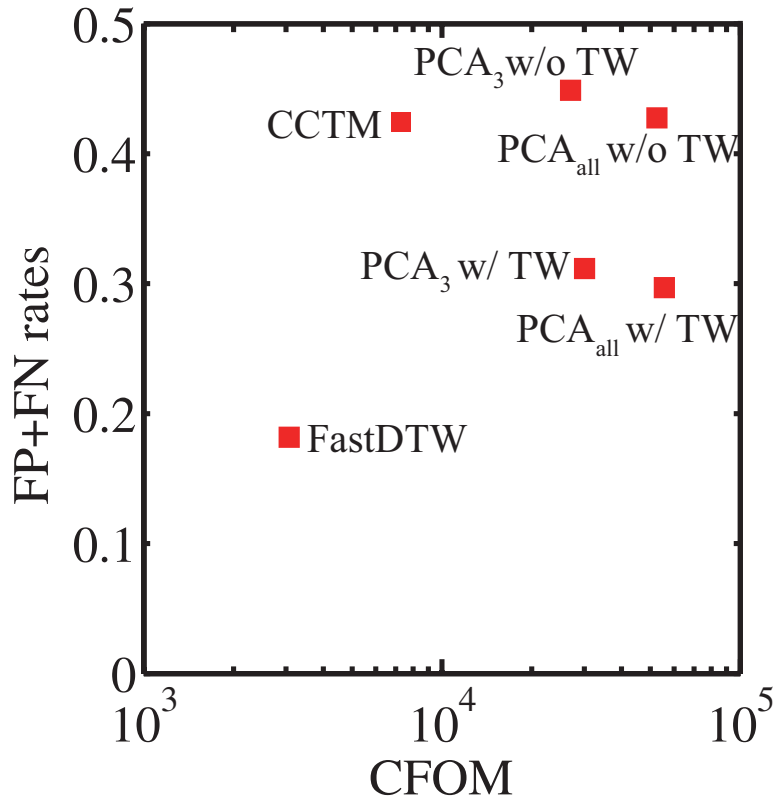


Figure 8.10: Trade-off between the FP+FN rates (averaged over all SNRs) and CFOM for different methods. Spike duration $\propto \text{norm}(1,0.3)$ in the synthesized recording.

Furthermore, as can be seen in Table 8.1, fastDTW only requires addition, condition and right/left shift to run, in contrast to PCA + k-means clustering

requiring multiplication/division and square root. The operations needed by fastDTW can potentially be implemented more efficiently in hardware (just like matched filter can be used for CCTM), so that the recording in each channel is processed in parallel with less than $2 \mu\text{W}/\text{recording channel}$.

Table 8.1: Complexity per spike of different methods.

Method	Conditions	Additions	Multiplications	Divisions	shifts	Square root
PCA+k-means w/o TW	kir	$2n^2 + 2n + kpir$	$4n^2 + 2n + kpir$	pir	-	kir
PCA+k-means w/ TW	$kir + 8kn$	$2n^2 + (8k + 2)n + kpir$	$4n^2 + 2n + kpir$	pir	$8kn$	kir
FastDTW	$8kn$	$8kn$	-	-	$8kn$	-
CCTM	-	knm	knm	-	-	-

Table 8.2: Variables to estimate the computational complexity.

Variable Symbol	Description	Value
n	number of samples in each spike	64
k	number of templates (clusters)	2
p	numebr of features per spike, $p \leq n$	3
i	number of iterations to find the local optimum for k-means clustering	10
r	number of replicates so that the local optimum equals to the global optimum for k-means clustering	3
m	number of stretched (shrunk) and shifted segments for each candidate spike for CCTM	3×10

CHAPTER 9

EXPERIMENTAL SPIKE CLASSIFICATION BY FASTDTW

In the previous chapter, we introduced fastDTW algorithm and demonstrated its superiority in classification accuracy and computational complexity. In this chapter we will apply fastDTW to experimental mouse enteric neural recordings. The recordings were performed with different stimuli and inhibitions with an emphasis on how variability can be captured in the presence of high noise level. Enabled by the more precise spike classification by fastDTW, statistical analysis on the time evolution of waveforms under different conditions were performed for functional study. The experimental setup for EAP recordings is described in Section (9.1). The experimental recordings and the classified waveforms were analyzed in Section (9.2). The correlation between the various nonstationary effects were further discussed in Section(9.3) to shed light on neural activities in the mouse ENS.

9.1 Experimental methods

We used a non-amperimetric C_vMOS sensor ([45]) for the experimental EAP recording of mouse ENS neurons in the jejunum tissues. As large ambient variability is expected from motility and frequent vesicle release, C_vMOS can use its control gate to pin the transistor operating point for less waveform distortion and reduce reliance on the large reference electrode which is impractical for future *in vivo* monitoring. The operating principle of the C_vMOS sensor was described in the supplementary material. Intestine tissues were harvested from C57BL/6J mice (Jackson Laboratory) about 20 minutes before the recording. Longitudinal muscle-myenteric plexus (LMMP) tissue was isolated and kept

in phosphate buffered saline (PBS) of pH 7.4 at room temperature. A $1\text{mm} \times 1\text{mm}$ section of LMMP was placed onto the sensing electrode chip with $200\mu\text{m} \times 200\mu\text{m}$ gold electrodes, which was secured by a 3D-printed fluidic well of epoxy. The electrode chip was wire-bonded to the sensing gate of the CvMOS sensor as in Fig. 9.1. The neural activities were then captured by the output drain current of CvMOS. A Keithley 2400 was used to set the transistor operating point. The drain current was converted to voltage by the transimpedance amplifier (TIA) (Stanford Research System SR570, CA, USA) with a sensitivity of $100\mu\text{A}/\text{V}$ and a band-pass filtering from 1 Hz to 300 kHz. The TIA output was collected on the computer through a data acquisition test board (NI BNC 2110 and NI USB 6259). The tissue was chemically (300mM KCl) or mechanically (tweezers stretching) stimulated to induce action potentials. To verify the action potential signal origin, the LMMP neural signal was later suppressed through the addition of 1mM tetrodotoxin (TTX).

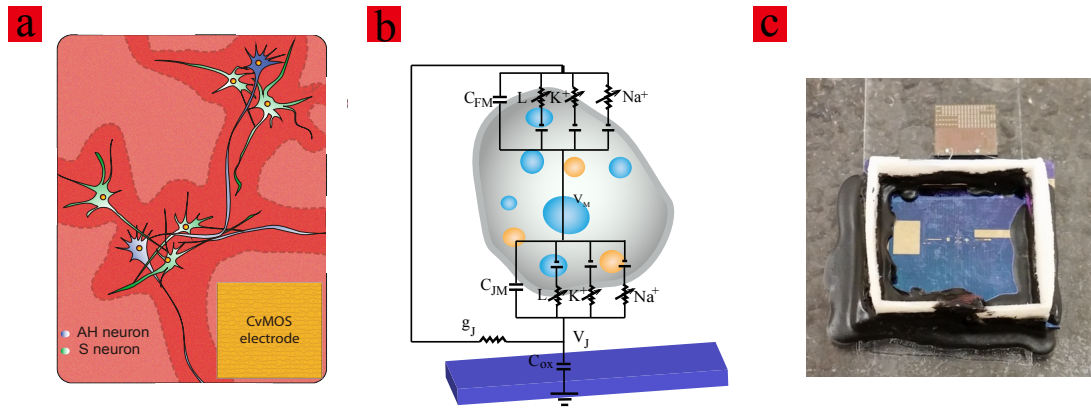


Figure 9.1: (a) Morphology of the neurons in mouse ENS. (b) A two-compartment circuit model of the neuron and CvMOS sensor interface. (c) A picture of the CvMOS sensor (the small chip on glass), and the electrode chip (the large chip with the fluidic well) used for the EAP recording.

9.2 Variable spike waveforms of experiment data

Mouse ENS EAP recording with high variability from cellular motility and vesicle release was performed on a non-amperimetric C_vMOS sensor with stablized biasing [45] and analyzed by fastDTW and CCTM. The segment width is chosen to be 6.4ms to allow for large variation in spike widths. The time bin to learn the average spike templates in the initialization phase is 1s and is changed to 0.5s in the later classification phase.

Two major spike waveforms were recognized: the biphasic (BP) and the monophasic (MP) spikes. EAP waveforms reflect the time profiles of the total membrane current as the summation of sodium, potassium, capacitive and leakage currents ([32]). The magnitude and the time course are further determined by the neuron type and morphology.

The bandpassed experimental recordings are plotted for mechanical stimulation (Fig. 9.2(a) top), chemical stimulation (Fig. 9.2(b) top), and TTX inhibition (Fig. 9.2(c) top). Mechanical stimulation by tissue stretching was applied twice at about 1s and 35s, which mimics stretching by intestinal peristalsis and stimulates serotonin-mediated enteric neuron firing. The peaks with magnitude over 3mV (Fig. 9.2(a) top around 40s) are most likely caused by movements between tissues and electrodes during the tweezers touching and stretching procedure. The captured BP and MP spikes are zoomed in during a 0.5s segment under the mechanical stimulation, which has magnitude below 0.25 mV (Fig. 9.2(d)). Chemical stimulation and inhibition was performed by KCl (Fig. 9.2(b)) and TTX (Fig. 9.2(c)) addition about 5 minutes before the recording, limited by diffusion.

The firing rates sorted by fastDTW over time are shown in Fig. 9.2 2nd row. The firing rate of BP spikes increased immediately after the mechanical stimulation applied at 35s. Both the BP and MP spikes ebbed away gradually after 90s. The magnitude (Fig. 9.2(a) 4th row) of the captured spikes peaked immediately after 35s, and returned to a smaller value after 50s. This suggests that mechanical stimulation increased enteric neuron firing, which subsided after a period. In contrast to the mechanical stimulation, neuronal firing stimulated by chemical stimulation sustained much longer, and the magnitude remained almost unchanged. TTX shuts down the sodium channel, and accordingly we observed the BP spike stopped, followed by the MP spikes (Only first 10s of the 120s long recording shown, no spike was present after the first 10s).

In comparison with CCTM (Fig. 9.2, 3rd row), fastDTW can not only count and capture firing more accurately, but also report the waveform evolution over time. The half-widths of the first and second phases of the BP spikes, and that of the MP spikes before, during and after the mechanical stimulation at 35s are depicted in Fig.9.3(a). Comparing before and right after stimulation, the half-width of the 1st phase of the BP spikes increased from about 0.75 ms to 1 ms. The half-widths of both the 2nd phase of the BP spikes (0.71 ms) and the MP spikes (0.53 ms) remained relatively constant. These waveform features are also demonstrated by the re-aligned median waveforms over time (Fig. 9.3(b)(c)), with clearly observable variations. For chemical stimulation, the BP spikes demonstrated a more constant 1st phase half-width of about 1 ms (Fig. 9.4(a)). The half-width of the 2nd phase of the BP spikes, as well as that of the MP spikes, decreased gradually from 0.9 ms to 0.63 ms, and from 0.79 ms to 0.6 ms, respectively. The re-aligned waveforms are shown in Figs. 9.4(b)(c).

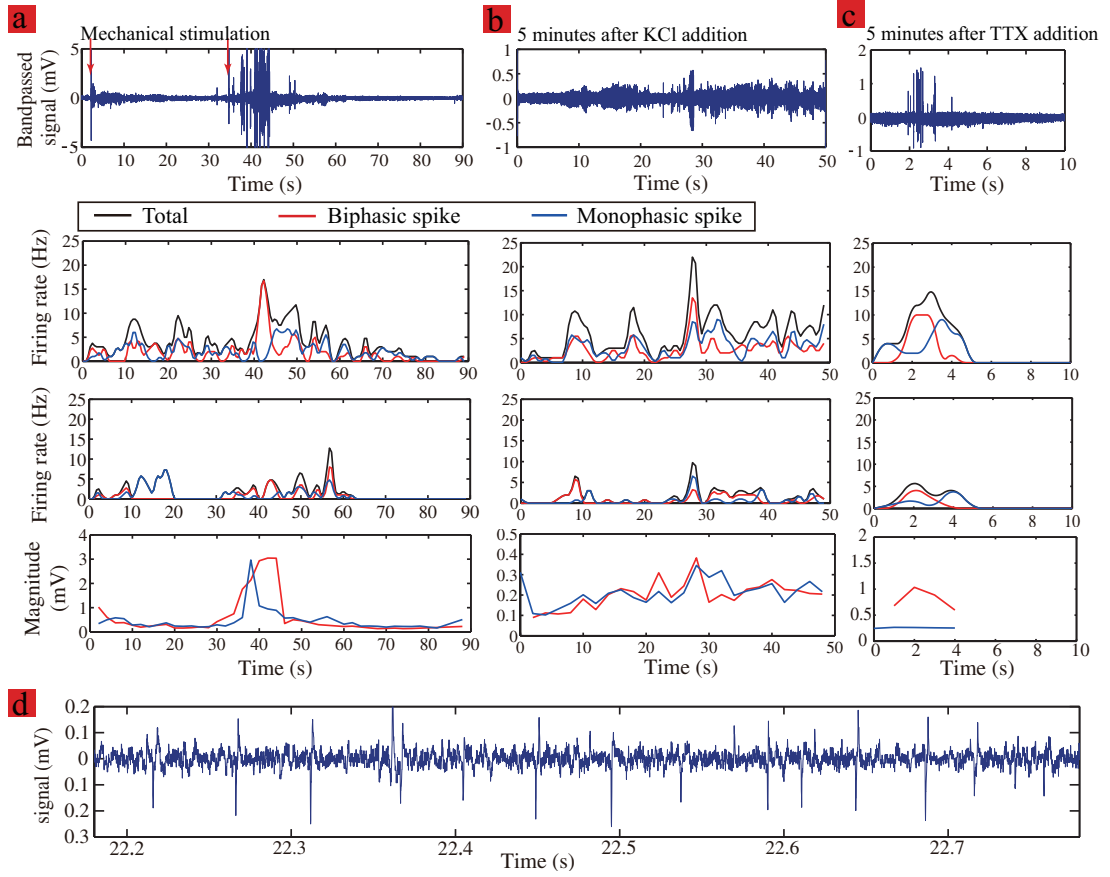


Figure 9.2: Bandpassed experimental recording (1st row), firing rate of spikes classified by fastDTW (2nd row) and by CCTM (3rd row), averaged spike magnitude (4th row) over time for (a) mechanical stimulation, (b) chemical stimulation, and (c) TTX inhibition. (d) A zoom-in view of the 0.5-second segment under the mechanical stimulation, showing the spike train. Notice that the waveform features with magnitude larger than 3mV under mechanical stimulation around 40 s were not recognized as action potentials, but more likely due to movement between tissues and electrodes. Due to the large variability in ENS EAP recording, the robust classification algorithm is very critical, because user inspection can be misleading.

FastDTW captures approximately 3 times more spikes than CCTM overall, which can cause inaccurate interpretation of ENS neural responses. The additional spike waveforms associated with larger variability recognized by fastDTW provide new insights to understand the ENS. The enhanced accuracy and efficiency as well as the detailed waveform reports in fastDTW offer a useful tool

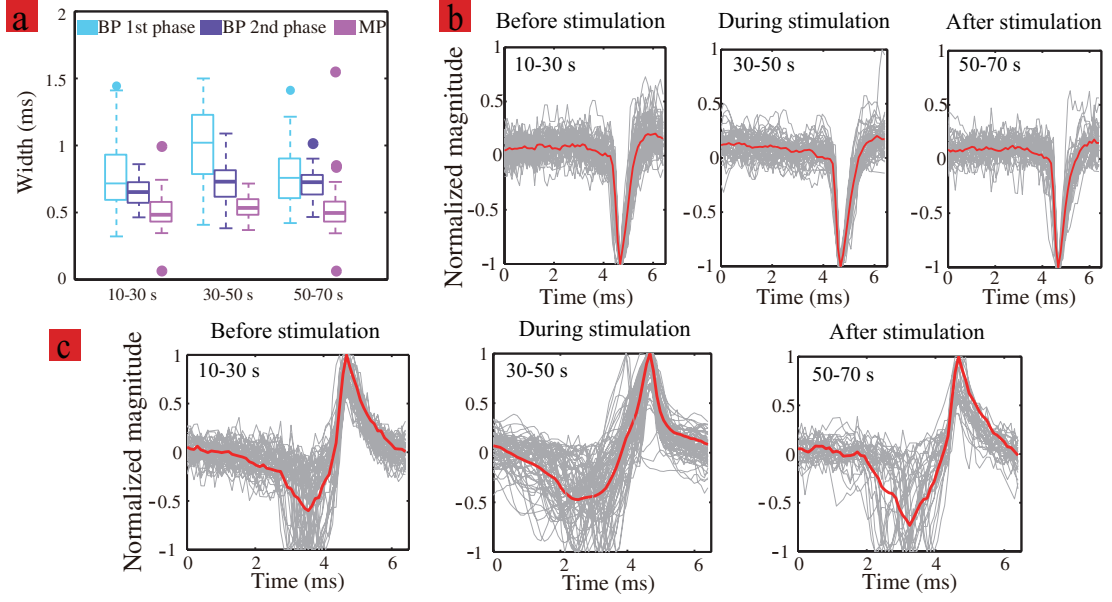


Figure 9.3: (a) The Box-and-whisker plot for the half-widths of the mechanically induced spikes; The aligned waveforms of (b) MP and (c) BP spikes before, during and after the mechanical stimulation. The red lines represent the medium waveforms. The nonstationary waveform features can be clearly observed.

to study the not-well understood and highly variable ENS from EAP recording.

The experimental data processed by PCA + k-means clustering can be found in the Supplementary Materials.

9.3 Discission on the nonstationary effects

To illustrate the new analyses enabled by fastDTW, we further studied the correlation between the various nonstationary effects in the mouse ENS experiments.

When the captured spikes are sorted by their firing rate, for both mechanical and chemical stimulation, the BP spikes have a slightly longer half-width in the 1st phase than in the 2nd phase (Fig. 9.5). An increase of the BP spike half-width at firing rates above 10 Hz was observed for mechanical stimulation, while the

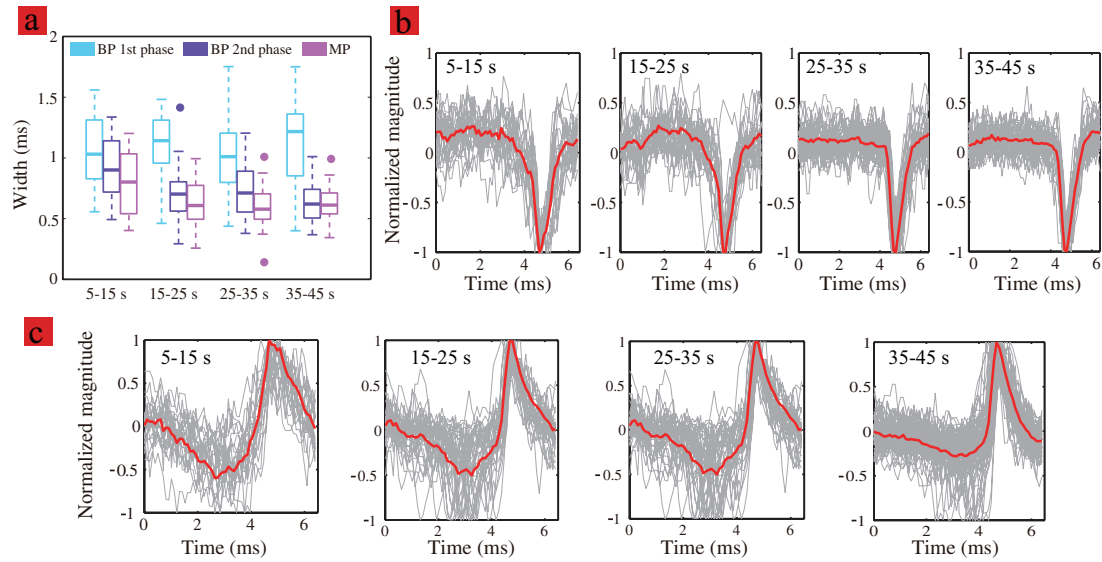


Figure 9.4: (a) The Box-and-Whisker plot for the half-width of the chemically induced spikes. The averaged waveforms over time, for the MP spikes (b) and BP spikes (c).

BP spike half-width shows no correlation with the firing rate for chemical stimulation. The magnitude of the BP spikes increases with the firing rate for both

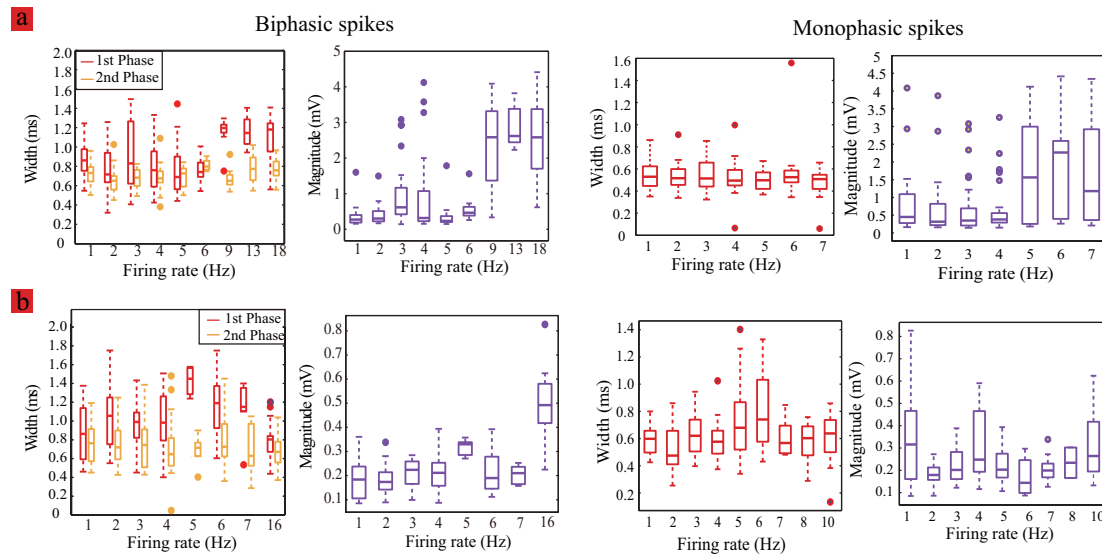


Figure 9.5: The Box-and-whisker plots of spike half-width and magnitude vs. the firing rate, (a) for mechanically induced spikes, and (b) for chemically induced spikes.

the mechanical and chemical stimulation. This may result from the occurrence of compound action potentials, when the neuron colony is highly excited. Another possibility is that spikes with higher magnitude and hence SNR are easier to sort out. The MP spikes showed no significant change in the half-width, and a slight increase of magnitude with the firing rate for mechanical stimulation.

The correlation coefficient between the spike half-width and the corresponding magnitude, and the significant level of the correlation (p value) are listed in Table 9.1. The correlation coefficients are marked in green for significant correlation, and red otherwise. For mechanically stimulated BP spikes, the spike half-width is positively related to the corresponding peak magnitude, which is possibly due to compound action potentials at high firing rate larger than 10 Hz. The BP spikes induced by chemical stimulation showed a slight decrease in the width with increasing magnitude at higher firing rate, indicated by the negative correlation coefficient, which has been observed in different neuron systems ([104]). For MP spikes induced mechanically, no significant correlation between the half-width and magnitude was found by the large p value, while the half-width of chemically induced MP spikes is negatively correlated to the magnitude.

Table 9.1: Correlation between the spike magnitude and half-width for mechanical and chemical stimulation.

Correlation coefficient	Biphasic spikes		Monophasic spikes	number of spikes
	1st phase	2nd phase		
Magnitude, mechanical stimulation	0.3564 (<.00001)	0.2673 (0.0002)	0.1091 (0.1461)	n = 190, 179
Magnitude, chemical stimulation	-0.4681 (<.00001)	-0.2695 (0.0003)	-0.2232 (0.0043)	n=174, 162

9.4 Conclusion

We have introduced fastDTW as an automatic spike classification method specially suited for real-time EAP recording with high waveform variability in time and magnitude. FastDTW provides improvements in accuracy and computational cost in comparison with CCTM and PCA + k-means clustering without time warping. We apply fastDTW to the mouse ENS neurons with high cellular motility and frequent vesicle release, using the CvMOS sensor with stabilized biasing in response to various stimuli. Biphasic and monophasic spikes are successfully recognized when the variability is as large as 1.2 ms in width and a few milli-volt in magnitude. We then illustrate how the captured waveform features can be used for variation correlation analyses. As fastDTW offers improved efficiency, accuracy and variability feature extraction, we believe it would benefit the ENS research community when large variability is expected in the EAP recording.

CHAPTER 10

CONCLUSION

10.1 Summary of major contributions

In this dissertation we present possible methods to deal with the variabilities in CMOS sensing networks. Specifically, we demonstrated device compensation based operational feedback to improve the performance of RF-to-DC rectifiers impacted by process variation, and developed signal processing algorithms based on time warping to enhance enteric neural recording and recognition in high noise high variability environment.

10.1.1 Contributions in RF-to-DC rectifier design

For the RF-to-DC rectifier design, our major contributions include:

(1) We designed the device structure of the tunable- V_{th} diode and developed SPICE circuit model for its simulation and design parameter optimization for further integration in the rectifier.

(2) We designed and fabricated the tunable- V_{th} rectifier based on the new diode structure, which to the best of our knowledge is the first demonstration of a rectifier where each diode can be tuned individually for better sensitivity and PCE.

(3) We developed an operational optimization algorithm, based on the feedback of the rectifier output, to program the diodes to optimal states. Because the

optimization happens experimentally on the fabricated chip, it can effectively overcome the process variation, which degrades the rectifier performance. Furthermore, the tuning algorithm allows each diode to be optimized for either low V_{th} or low leakage based on its role in the rectifier, which can greatly improve the overall system performance.

(4) We performed experimental characterization of the proposed RF-to-DC tunable rectifier, in comparison to a zero- V_{th} rectifier. The tunable- V_{th} rectifier can achieve -27 dBm sensitivity and 22% PCE at 547 MHz operating frequency, when used in combination with passive input boosting from matching network with $Q = 10$.

10.1.2 Contributions to enteric recording in high noise environment

For the neural recording platform in the enteric nervous system, our contributions include:

(1) We developed fastDTW as an automatic spike classification method specially suited for real-time EAP recording with high waveform variability in time and magnitude. FastDTW provides significant improvements in accuracy and computational cost in comparison with CCTM and PCA + k-means clustering without time warping.

(2) We applied fastDTW to the mouse ENS recordings with high cellular motility and frequent vesicle release, using the CMOS sensor with stabilized biasing in response to various stimuli. The proposed fastDTW spike classifica-

tion method successfully recognized biphasic and monophasic spikes when the variability is as large as 1.2 ms in width and a few milli-volt in magnitude.

(3) We illustrated how the captured waveform features can be used for variation correlation analyses to shed light onto the working principle of the neural networks.

10.2 Suggestions for future work

There are multiple interesting and important directions that we can continue to explore in future works.

10.2.1 Future works on the RF smart sensing platform

For the RF smart sensing platform, future works can be done in the following areas.

Cross-coupled RF-to-DC rectifier based on tunable diodes

The sensitivity and PCE are two key metrics that determine the performance of the RF-to-DC rectifier, which is a critical component for systems that are powered by RF energy harvesting. Dickson and cross-coupled architectures are two common designs for the RF-to-DC rectifiers (Sec.5.4). Generally speaking, the Dickson type architectures demonstrate a better sensitivity, whereas the cross-coupled designs provide higher PCE. So Dickson type design is widely used for

applications where the system base on wake-up circuitry but has high requirement for the operating range, while the cross-coupled designs fit systems that rely on continuous energy harvesting.

But is there a rectifier design that can achieve better sensitivity without losing the PCE? In this dissertation we have demonstrated the ability of the tunable threshold diodes to improve the sensitivity in Dickson rectifiers, by optimizing the operating point of each diode for either low turn-on voltage or low leakage. The tunable threshold diodes can also be used in cross-coupled rectifiers to potentially enhance the sensitivity.

A preliminary exploration of the cross-coupled RF-to-DC rectifier based on tunable diodes was conducted in Section 6, where a 3-stage tunable rectifier and a 1-stage conventional one were implemented in Cadence. The simulation results show that a 3 dB improvement in the sensitivity can be achieved by the tunable cross-coupled rectifier over the conventional one, with comparable PCE values. Despite the slight improvement in the sensitivity observed in our simulation, there are more design trade-offs that needs to be carefully examined and dealt with for a complete study:

(1) For the same $V_{in,peak}$, bigger stage number leads to higher V_{out} , but it reduces PCE.

(2) large input impedance of the rectifier is needed for passive input voltage boosting in the matching. The input impedance decreases with stage numbers (because the stages are in parallel).

(3) The increase in the rectifier input impedance comes at a cost of more area and more charging time. And the input impedance can only be increased to a

certain extent.

In order to design a RF-to-DC rectifier with even better sensitivity and PCE, a more complete research work that includes the antenna, matching network and rectifier co-design is required. The design may need to go through a couple rounds of iterations and simulations to obtain the best performance.

Adaptive RF-to-DC rectifier

So far in the rectifier designs we have used the same dimension for all the transistors and all the capacitors. However in the simulation of a single stage at different positions, we have shown that different stages can have quite distinctive impact and functions on the rectifier. Therefore, it is possible that adaptive sizing of the devices can lead to further enhancement in sensitivity or PCE. Future work can be done to investigate the effect of adaptive transistor and capacitor sizing on the rectifier performance. Now that we have developed an testing flow in Cadence Virtuoso for automatic design parameter optimization, an algorithm can potentially be designed to guide the system design. The algorithm development itself is an interesting mathematical optimization problem that may offer insight into the rectifier design from an unique angle.

RFID systems based on the tunable- V_{th} rectifier

RF-to-DC rectifiers are crucial for the RFID systems. As shown in the previous sections, Dickson type rectifiers offer a good sensitivity, yet they rely zero- V_{th} transistors in special fabrication process that are difficult, and are susceptible to unavoidable process variations, especially in the continuously scaling technolo-

gies. Cross-coupled rectifiers provide an enhanced PCE without the need for special transistors, but are relatively inferior in terms of sensitivity. The tunable- V_{th} rectifier proposed in this dissertation can overcome the impact of process variation, and achieve better performance by tuning each individual diodes to either low V_{th} or low leakage operating points. An improved -27 dBm sensitivity together with 22% PCE was demonstrated at 547 MHz with a $Q = 10$ matching network.

Such improvement in the RF-to-DC rectifier make it possible for real-time RFID systems for a wide range of applications: such as RF indoor locating for object detection and item counting; real-time gesture/posture recognition for VR/AR; wireless recording platforms for bio-medical systems.

10.2.2 Future works on the spike sorting algorithm

Fast-DTW based algorithm for the recording of vital signals

In this dissertation, a spike classification algorithm based on fast-DTW has been developed with high accuracy in high noise enteric neural recordings. Although the algorithm was applied on enteric recordings in this work, it can be readily applied to other spike sorting problems that requires high tolerance to waveform variabilities in both the magnitude and the timing. For example, the algorithm can be used to pick up waveforms of heart-beat, respiration in vital signal recording systems [41]. The progresses in such systems can offer low-cost non-intrusive health monitoring options that can benefit millions of people with concerns on heart disease, sleeping disorder and respiration problems.

Real-time on-chip neural recording systems

The fast-DTW spike sorting algorithm shows an improved linear complexity compared to the traditional clustering based method. So the entire recording and processing system can be implemented on-chip. This on-chip implementation in regular CMOS process can greatly reduce the cost of the system. Its miniature size also make it less-intrusive for the biological neuron networks.

BIBLIOGRAPHY

- [1] ALC-370 Data Sheet: Higgstm-4 EPC Class 1 Gen 2 RFID Tag IC.
- [2] Mohamed A Abouzied and Edgar Sánchez-Sinencio. Low-input power-level cmos rf energy-harvesting front end. *IEEE Trans. on Microwave Theory and Techniques*, 63(11):3794–3805, 2015.
- [3] Dimitrios A Adamos, Efstratios K Kosmidis, and George Theophilidis. Performance evaluation of pca-based spike sorting algorithms. *Computer methods and programs in biomedicine*, 91(3):232–244, 2008.
- [4] Lesley Adkins, Roy A Adkins, et al. *Handbook to life in ancient Rome*. Infobase publishing, 2014.
- [5] Abdullah S. Almansouri, Mahmoud H. Ouda, and Khaled N. Salama. A CMOS RF-to-DC power converter with 86% efficiency and -19.2-dBm sensitivity. *IEEE Transactions on Microwave Theory and Techniques*, 2:1–7, 2018.
- [6] Asen Asenov, Savas Kaya, and John H Davies. Intrinsic threshold voltage fluctuations in decanano mosfets due to local oxide thickness variations. *IEEE Transactions on electron devices*, 49(1):112–119, 2002.
- [7] Aharon Bar-Hillel, Adam Spiro, and Eran Stark. Spike sorting: Bayesian clustering of non-stationary data. *Journal of neuroscience methods*, 157(2):303–316, 2006.
- [8] Gabrio Bassotti, Elisabetta Antonelli, Vincenzo Villanacci, Monia Baldoni, and Maria Pina Dore. Colonic motility in ulcerative colitis. *United European gastroenterology journal*, page 2050640614548096, 2014.
- [9] Mark Bohr. The evolution of scaling from the homogeneous era to the heterogeneous era. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 1–1. IEEE, 2011.
- [10] Miodrag Bolic, David Simplot-Ryl, and Ivan Stojmenovic. *RFID Systems: Research Trends and Challenges*. John Wiley & Sons, 2010.
- [11] JR Brews. Surface potential fluctuations generated by interface charge inhomogeneities in mos devices. *Journal of Applied Physics*, 43(5):2306–2313, 1972.

- [12] David Brooks and John Sartori. Ultra-low-power processors. *IEEE Micro*, 37(6):16–19, 2017.
- [13] Ana Calabrese and Liam Paninski. Kalman filter mixture model for spike sorting of non-stationary data. *Journal of neuroscience methods*, 196(1):159–169, 2011.
- [14] Olivier Carnal and Jürgen Mlynek. Youngs double-slit experiment with atoms: A simple atom interferometer. *Physical review letters*, 66(21):2689, 1991.
- [15] Davide Castelvecchi. Mysteries of turbulence unravelled. *Nature News*, 548(7668):382, 2017.
- [16] Hsiao-Lung Chan, Tony Wu, Shih-Tseng Lee, Shih-Chin Fang, Pei-Kuang Chao, and Ming-An Lin. Classification of neuronal spikes over the re-constructed phase space. *Journal of neuroscience methods*, 168(1):203–211, 2008.
- [17] Taiyun Chi, Hechen Wang, Min-Yu Huang, Fa Foster Dai, and Hua Wang. A bidirectional lens-free digital-bits-in/-out 0.57 mm 2 terahertz nano-radio in cmos with 49.3 mw peak power consumption supporting 50cm internet-of-things communication. In *Custom Integrated Circuits Conference (CICC), 2017 IEEE*, pages 1–4. IEEE, 2017.
- [18] Wim Cornelissen, Ann De Laet, Alfons BA Kroese, Pierre-Paul Van Bogaert, Dietrich W Scheuermann, and Jean-Pierre Timmermans. Electrophysiological features of morphological dogiel type ii neurons in the myenteric plexus of pig small intestine. *Journal of neurophysiology*, 84(1):102–111, 2000.
- [19] Jari Pascal Curty, Norbert Joehl, Catherine Dehollain, and Michel J. Declercq. Remotely powered addressable UHF RFID integrated system. *IEEE J. of Solid-State Circuits*, 40(11):2193–2202, 2005.
- [20] Roberto De Giorgio, Stefania Guerrini, Giovanni Barbara, Vincenzo Stanghellini, Fabrizio De Ponti, Roberto Corinaldesi, Peter L Moses, Keith A Sharkey, and Gary M Mawe. Inflammatory neuropathies of the enteric nervous system. *Gastroenterology*, 126(7):1872–1883, 2004.
- [21] J.F. Dickson. On-chip high-voltage generation in MNOS integrated circuits using an improved voltage multiplier technique. *IEEE J. of Solid-State Circuits*, 11(3):374–378, 1976.

- [22] Alan Drake, Robert Senger, Harmander Deogun, Gary Carpenter, Soraya Ghiasi, Tuyet Nguyen, Norman James, Michael Floyd, and Vikas Pokala. A distributed critical-path timing monitor for a 65nm high-performance microprocessor. In *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, pages 398–399. IEEE, 2007.
- [23] Max Eickenscheidt, Martin Jenkner, Roland Thewes, Peter Fromherz, and Günther Zeck. Electrical stimulation of retinal neurons in epiretinal and subretinal configuration using a multicapacitor array. *Journal of neurophysiology*, 107(10):2742–2755, 2012.
- [24] Chaitanya Ekanadham, Daniel Tranchina, and Eero P Simoncelli. A unified framework and method for automatic neural spike identification. *Journal of neuroscience methods*, 222:47–55, 2014.
- [25] Felix Franke, Rodrigo Quian Quiroga, Andreas Hierlemann, and Klaus Obermayer. Bayes optimal template matching for spike sorting—combining fisher discriminant analysis with optimal filtering. *Journal of computational neuroscience*, 38(3):439–459, 2015.
- [26] Peter Fromherz. Extracellular recording with transistors and the distribution of ionic conductances in a cell membrane. *European Biophysics Journal*, 28(3):254–258, 1999.
- [27] Hidenobu Fukutome, Youichi Momiyama, Tomohiro Kubo, Yukio Tagawa, Takayuki Aoyama, and Hiroshi Arimoto. Direct evaluation of gate line edge roughness impact on extension profiles in sub-50-nm n-mosfets. *IEEE Transactions on Electron Devices*, 53(11):2755–2763, 2006.
- [28] John B Furness, Brid P Callaghan, Leni R Rivera, and Hyun-Jung Cho. The enteric nervous system and gastrointestinal innervation: integrated local and central control. In *Microbial endocrinology: The microbiota-gut-brain axis in health and disease*, pages 39–71. Springer, 2014.
- [29] Aleena Garner and Mark Mayford. New approaches to neural circuits in behavior. *Learning & Memory*, 19(9):385–390, 2012.
- [30] Kaveh Gharehbaghi, Özge Zorlu, Fatih Koçer, and Haluk Külâh. Auto-calibrating threshold compensation technique for rf energy harvesters. In *Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 179–182. IEEE, 2015.

- [31] Sarah Gibson, Jack W Judy, and Dejan Marković. Comparison of spike-sorting algorithms for future hardware implementation. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 5015–5020. IEEE, 2008.
- [32] Carl Gold, Darrell A Henze, Christof Koch, and György Buzsáki. On the origin of the extracellular action potential waveform: a modeling study. *Journal of neurophysiology*, 95(5):3113–3128, 2006.
- [33] Anders Hald. *A history of probability and statistics and their applications before 1750*, volume 501. John Wiley & Sons, 2003.
- [34] Benedikt Hallgrímsson and Brian K Hall. Variation and variability: central concepts in biology. In *Variation*, pages 1–7. Elsevier, 2005.
- [35] Zohaib Hameed and Kambiz Moez. Hybrid forward and backward threshold-compensated RF-DC power converter for RF energy harvesting. *IEEE J. Emerging and Selected Topics in Circuits and Systems*, 4(3):335–343, 2014.
- [36] Zohaib Hameed and Kambiz Moez. A 3.2 V -15 dBm Adaptive Threshold-Voltage Compensated RF Energy Harvester in 130 nm CMOS. *IEEE Trans. on Circuits and Systems I: Regular Papers*, 62(4):948–956, 2015.
- [37] Owen P Hamill, A Marty, Erwin Neher, Bert Sakmann, and FJ Sigworth. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv*, 391(2):85–100, 1981.
- [38] Osamu Hirabayashi, Atsushi Kawasumi, Azuma Suzuki, Yasuhisa Takeyama, Keiichi Kushida, Takahiko Sasaki, Akira Katayama, Gou Fukano, Yuki Fujimura, Takaaki Nakazato, et al. A process-variation-tolerant dual-power-supply sram with 0.179 μm 2 cell in 40nm cmos using level-programmable wordline driver. In *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 458–459. IEEE, 2009.
- [39] Alston S Householder. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM (JACM)*, 5(4):339–342, 1958.
- [40] Bin Huang and W Kinsner. Ecg frame classification using dynamic time warping. In *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*, volume 2, pages 1105–1110. IEEE, 2002.

- [41] Xiaonan Hui and Edwin C Kan. Monitoring vital signs over multiplexed radio by near-field coherent sensing. *Nature Electronics*, 1(1):74, 2018.
- [42] Jan D Huizinga and Ji-Hong Chen. The myogenic and neurogenic components of the rhythmic segmentation motor patterns of the intestine. *Frontiers in neuroscience*, 8, 2014.
- [43] Michael Hutzler, Armin Lambacher, Bjoern Eversmann, Martin Jenkner, Roland Thewes, and Peter Fromherz. High-resolution multitransistor array recording of electrical field potentials in cultured brain slices. *Journal of neurophysiology*, 96(3):1638–1645, 2006.
- [44] David Jäckel, Urs Frey, Michele Fiscella, Felix Franke, and Andreas Hierlemann. Applicability of independent component analysis on high-density microelectrode array recordings. *Journal of neurophysiology*, 108(1):334–348, 2012.
- [45] Krishna Jayant, Kshitij Auluck, Mary Funke, Sharlin Anwar, Joshua B Phelps, Philip H Gordon, Shantanu R Rajwade, and Edwin C Kan. Programmable ion-sensitive transistor interfaces. i. electrochemical gating. *Physical Review E*, 88(1):012801, 2013.
- [46] Jian Kang, Patrick Yin Chiang, and Arun Natarajan. 21.6 a 1.2 cm² 2.4 GHz self-oscillating rectifier-antenna achieving -34.5 dBm sensitivity for wirelessly powered sensors. In *Solid-State Circuits Conference (ISSCC)*, pages 374–375. IEEE, 2016.
- [47] Vaibhav Karkare, Sarah Gibson, and Dejan Markovic. A 75- μ w, 16-channel neural spike-sorting processor with unsupervised clustering. *Solid-State Circuits, IEEE Journal of*, 48(9):2230–2238, 2013.
- [48] Sunghan Kim and James McNames. Automatic spike detection based on adaptive template matching for extracellular neural recordings. *Journal of neuroscience methods*, 165(2):165–174, 2007.
- [49] Youchang Kim, Dongjoo Shin, Jinsu Lee, Yongsu Lee, and Hoi-Jun Yoo. A 0.55 v 1.1 mw artificial intelligence processor with on-chip pvt compensation for autonomous mobile robots. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(2):567–580, 2018.
- [50] Kazuo Kitamura, Benjamin Judkewitz, Masanobu Kano, Winfried Denk, and Michael Häusser. Targeted patch-clamp recordings and single-cell

- electroporation of unlabeled neurons in vivo. *Nature methods*, 5(1):61, 2008.
- [51] Fiath Kocer and Michael P Flynn. A new transponder architecture for long-range telemetry applications. *Proc. of the 2005 European Conference on Circuit Theory and Design*, 2(5):177–180, 2005.
 - [52] Suhasa B Kodandaramaiah, Giovanni Talei Franzesi, Brian Y Chow, Edward S Boyden, and Craig R Forest. Automated whole-cell patch-clamp electrophysiology of neurons in vivo. *Nature methods*, 9(6):585, 2012.
 - [53] Koji Kotani, Atsushi Sasaki, and Takashi Ito. High-efficiency differential-drive cmos rectifier for uhf rfids. *IEEE Journal of Solid-State Circuits*, 44(11):3011–3018, 2009.
 - [54] Kelin J Kuhn. Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale cmos. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 471–474. IEEE, 2007.
 - [55] T Le, K Mayaram, and T Fiez. Efficient far-field radio frequency energy harvesting for passive powered sensor networks. *IEEE Journal of Solid-State Circuits*, 43(5):1287–1302, 2008.
 - [56] Triet Le, Karti Mayaram, and Terri Fiez. Efficient far-field radio frequency energy harvesting for passively powered sensor networks. *IEEE J. of Solid-State Circuits*, 43(5):1287–1302, 2008.
 - [57] Mikhail A Lebedev and Miguel AL Nicolelis. Brain-machine interfaces: past, present and future. *TRENDS in Neurosciences*, 29(9):536–546, 2006.
 - [58] Hyung Min Lee and Maysam Ghovanloo. An integrated power-efficient active rectifier with offset-controlled high speed comparators for inductively powered applications. *IEEE Trans. on Circuits and Systems I: Regular Papers*, 58(8):1749–1760, 2011.
 - [59] Michael S Lewicki. Bayesian modeling and classification of neural signals. *Neural computation*, 6(5):1005–1030, 1994.
 - [60] Bo Li, Xi Shao, Negin Shahshahan, Neil Goldsman, Thomas Salter, and George M Metze. An antenna co-design dual band rf energy harvester.

IEEE Transactions on Circuits and Systems I: Regular Papers, 60(12):3256–3266, 2013.

- [61] Stuart P Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.
- [62] Alan E Lomax, David R Linden, Gary M Mawe, and Keith A Sharkey. Effects of gastrointestinal inflammation on enteroendocrine cells and enteric neural reflex circuits. *Autonomic Neuroscience*, 126:250–257, 2006.
- [63] Carolina Mora Lopez, Dimiter Prodanov, Dries Braeken, Ivan Gligorijevic, Wolfgang Eberle, Carmen Bartic, Robert Puers, and Georges Gielen. A multichannel integrated circuit for electrical recording of neural activity, with independent channel programmability. *Biomedical Circuits and Systems, IEEE Transactions on*, 6(2):101–110, 2012.
- [64] Yan Lu, Haojuan Dai, Mo Huang, Man Kay Law, Sai Weng Sin, U. Seng-Pan, and Rui P. Martins. A wide input range dual-path CMOS rectifier for RF energy harvesting. *IEEE Trans. on Circuits and Systems II: Express Briefs*, 64(2):166–170, 2017.
- [65] Ronald Luijten, Dae Pham, Rolf Clauberg, Matteo Cossale, Huy N Nguyen, and Mihir Pandya. 4.4 energy-efficient microserver based on a 12-core 1.8 ghz 188k-coremark 28nm bulk cmos 64b soc for big-data applications with 159gb/s/l memory bandwidth system density. In *Solid-State Circuits Conference-(ISSCC), 2015 IEEE International*, pages 1–3. IEEE, 2015.
- [66] Yanjun Ma and Edwin C. Kan. *Non-logic Devices in Logic Processes*. Springer, 2017.
- [67] Yunfei Ma, Xiaonan Hui, and Edwin C Kan. 3D real-time indoor localization via broadband nonlinear backscatter in passive devices with centimeter precision. In *Proc. of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 216–229. ACM, 2016.
- [68] Yunfei Ma, Zhihong Luo, Christoph Steiger, Giovanni Traverso, and Fadel Adib. Enabling Deep-Tissue Networking for Miniature Medical Devices. In *ACM SIGCOMM*, 2018.
- [69] Hamid Mahmoodi, Saibal Mukhopadhyay, and Kaushik Roy. Estimation of delay variations due to random-dopant fluctuations in nanoscale cmos circuits. *IEEE Journal of Solid-State Circuits*, 40(9):1787–1796, 2005.

- [70] Blake R. Marshall, Marcin M. Morys, and Gregory D. Durgin. Parametric analysis and design guidelines of RF-to-DC Dickson charge pumps for RFID energy harvesting. *IEEE International Conf. on RFID*, pages 32–39, 2015.
- [71] Deepak Mishra, Swades De, Soumya Jana, Stefano Basagni, Kaushik Chowdhury, and Wendi Heinzelman. Smart rf energy harvesting communications: Challenges and opportunities. *IEEE Communications Magazine*, 53(4):70–78, 2015.
- [72] Rachit Mohan, Samira Zaliasl, Georges GE Gielen, Chris Van Hoof, Refet Firat Yazicioglu, and Nick Van Helleputte. A 0.6-v, 0.015-mm², time-based ecg readout for ambulatory applications in 40-nm cmos. *IEEE Journal of Solid-State Circuits*, 52(1):298–308, 2017.
- [73] Saibal Mukhopadhyay, Keejong Kim, Hamid Mahmoodi, and Kaushik Roy. Design of a process variation tolerant self-repairing sram for yield enhancement in nanoscaled cmos. *IEEE Journal of Solid-State Circuits*, 42(6):1370–1382, 2007.
- [74] Jan Müller, Marco Ballini, Paolo Livi, Yihui Chen, Milos Radivojevic, Amir Shadmani, Vijay Viswam, Ian L Jones, Michele Fiscella, Roland Diggelmann, et al. High-resolution cmos mea platform to study neurons at sub-cellular, cellular, and network levels. *Lab on a Chip*, 15(13):2767–2780, 2015.
- [75] Joaquin Navajas, Deren Y Barsakcioglu, Amir Eftekhari, Andrew Jackson, Timothy G Constandinou, and Rodrigo Quian Quiroga. Minimum requirements for accurate and efficient real-time on-chip spike sorting. *Journal of neuroscience methods*, 230:51–64, 2014.
- [76] Thanh Nguyen, Abbas Khosravi, Douglas Creighton, and Saeid Naha-vandi. Spike sorting using locality preserving projection with gap statistics and landmark-based spectral clustering. *Journal of neuroscience methods*, 238:43–53, 2014.
- [77] Miguel AL Nicolelis. Brain-machine interfaces to restore motor function and probe neural circuits. *Nature Reviews Neuroscience*, 4(5):417, 2003.
- [78] Kulmira Nurgali, Martin J Stebbing, and John B Furness. Correlation of electrophysiological and morphological characteristics of enteric neurons in the mouse colon. *Journal of Comparative Neurology*, 468(1):112–124, 2004.

- [79] Seunghyun Oh, David D Wentzloff, and Ann Arbor. A -32 dBm sensitivity RF power harvester in 130 nm CMOS. *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, (2):483–486, 2012.
- [80] Katsuhiko Ohsaki, Noriaki Asamoto, and Shunichi Takagaki. A single poly eeprom cell structure for use in standard cmos processes. *IEEE J. of Solid-State Circuits*, 29(3):311–316, 1994.
- [81] Mahmoud H Ouda, Waleed Khalil, and Khaled N Salama. Self-biased differential rectifier with enhanced dynamic range for wireless powering. *IEEE Trans. on Circuits and Systems II: Express Briefs*, 64(5):515–519, 2017.
- [82] Sivylla E Paraskevopoulou, Deren Y Barsakcioglu, Mohammed R Saberi, Amir Eftekhari, and Timothy G Constandinou. Feature extraction using first and second derivative extrema (fsde) for real-time and hardware-efficient spike sorting. *Journal of neuroscience methods*, 215(1):29–37, 2013.
- [83] Sivylla E Paraskevopoulou, Di Wu, Amir Eftekhari, and Timothy G Constandinou. Hierarchical adaptive means (ham) clustering for hardware-efficient, unsupervised and real-time spike sorting. *Journal of neuroscience methods*, 235:145–156, 2014.
- [84] Mirko Pasca, Stefano D’Amico, Vincenzo Chironi, Luca Catarinucci, Danilo De Donno, Riccardo Colella, and Luciano Tarricone. A -19dBm sensitivity integrated RF-DC converter with regulated output voltage for powering UHF wireless sensors. *Proc. 6th IEEE International Workshop on Advances in Sensors and Interfaces, IWASI 2015*, pages 168–171, 2015.
- [85] Andrei Pavlov and Manoj Sachdev. *CMOS SRAM circuit design and parametric test in nano-scaled technologies: process-aware SRAM design and test*, volume 40. Springer Science & Business Media, 2008.
- [86] Jonathan W Pillow, Jonathon Shlens, EJ Chichilnisky, and Eero P Simoncelli. A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings. *PloS one*, 8(5):e62123, 2013.
- [87] Jason S Prentice, Jan Homann, Kristina D Simmons, Gašper Tkačik, Vijay Balasubramanian, and Philip C Nelson. Fast, scalable, bayesian spike identification for multi-electrode arrays. *PloS one*, 6(7):e19884, 2011.
- [88] Yu Pu, Chunlei Shi, Giby Samson, Dongkyu Park, Ken Easton, Rudy Beraha, Adam Newham, Mark Lin, Venkat Rangan, Karam Chatha, et al. A

- 9-mm² ultra-low-power highly integrated 28-nm cmos soc for internet of things. *IEEE Journal of Solid-State Circuits*, 2018.
- [89] R Quian Quiroga. What is the real shape of extracellular spikes? *Journal of neuroscience methods*, 177(1):194–198, 2009.
- [90] R Quian Quiroga, Zoltan Nadasdy, and Yoram Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*, 16(8):1661–1687, 2004.
- [91] Juan C. Ranuárez, M.J. Deen, and Chih-Hung Chen. A review of gate tunneling current in MOS devices. *Microelectronics Reliability*, 46(12):1939–1956, 2006.
- [92] John M Rist et al. *Epicurus: an introduction*. CUP Archive, 1972.
- [93] Michael Rizk, Chad A Bossetti, Thomas A Jochum, Stephen H Callender, Miguel AL Nicolelis, Dennis A Turner, and Patrick D Wolf. A fully implantable 96-channel neural data acquisition system. *Journal of neural engineering*, 6(2):026002, 2009.
- [94] Ueli Rutishauser, Erin M Schuman, and Adam N Mamelak. Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. *Journal of neuroscience methods*, 154(1):204–224, 2006.
- [95] Zahra Safarian and Hossein Hashemi. Wirelessly powered passive systems with dynamic energy storage mechanism. *IEEE Trans. on Microwave Theory and Techniques*, 62(4):1012–1021, 2014.
- [96] Stan Salvador and Philip Chan. Fastdtw: Toward accurate dynamic time warping in linear time and space. In *KDD workshop on mining temporal and sequential data*. Citeseer, 2004.
- [97] Tor C Savidge, Michael V Sofroniew, and Michel Neunlist. Starring roles for astroglia in barrier pathologies of gut and brain. *Laboratory investigation*, 87(8):731–736, 2007.
- [98] S. Scorcioni, L. Larcher, and A. Bertacchini. Optimized CMOS RF-DC converters for remote wireless powering of RFID applications. *IEEE International Conf. on RFID*, pages 47–53, 2012.

- [99] Bert Serneels, Michiel Steyaert, and Wim Dehaene. A high speed, low voltage to high voltage level shifter in standard 1.2 v 0.13 μm cmos. *Analog Integrated Circuits and Signal Processing*, 55(1):85–91, Apr. 2008.
- [100] Sajjad Shieh and Mahmoud Kamarei. Transient input impedance modeling of rectifiers for RF energy harvesting applications. *IEEE Trans. on Circuits and Systems II: Express Briefs*, 65(3):311–315, 2018.
- [101] M Sipser. Introduction to the theory of computation . course technology, 2005.
- [102] Nick J Spencer and Terence K Smith. Mechanosensory s-neurons rather than ah-neurons appear to generate a rhythmic motor pattern in guinea-pig distal colon. *The Journal of physiology*, 558(2):577–596, 2004.
- [103] Mark Stoopman, Student Member, and Shady Keyrouz. Co-design of a CMOS rectifier and small loop antenna for highly sensitive RF energy harvesters. *IEEE J. of Solid-State Circuits*, 49(3):622–634, 2014.
- [104] Peter Stratton, Allen Cheung, Janet Wiles, Eugene Kiyatkin, Pankaj Sah, and François Windels. Action potential waveform variability limits multi-unit separation in freely behaving rats. *PloS one*, 7(6):e38482–e38482, 2012.
- [105] RH Straub, R Wiest, UG Strauch, P Härle, and J Schölmerich. The role of the sympathetic nervous system in intestinal inflammation. *Gut*, 55(11):1640–1649, 2006.
- [106] Jack Y-C Sun. System scaling for intelligent ubiquitous computing. In *Electron Devices Meeting (IEDM), 2017 IEEE International*, pages 1–3. IEEE, 2017.
- [107] JH Szurszewski, LG Ermilov, and SM Miller. Prevertebral ganglia and intestinofugal afferent neurones. *Gut*, 51(suppl 1):i6–i10, 2002.
- [108] S Takahashi and Y Sakurai. Real-time and automatic sorting of multi-neuronal activity for sub-millisecond interactions in vivo. *Neuroscience*, 134(1):301–315, 2005.
- [109] Susumu Takahashi, Yuichiro Anzai, and Yoshio Sakurai. Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. *Journal of neurophysiology*, 89(4):2245–2258, 2003.

- [110] Yuan Taur. Cmos design near the limit of scaling. *IBM Journal of Research and Development*, 46(2.3):213–222, 2002.
- [111] Lars Thuneberg and Susan Peters. Toward a concept of stretch-coupling in smooth muscle. i. anatomy of intestinal segmentation and sleeve contractions. *The Anatomical Record*, 262(1):110–124, 2001.
- [112] T. Umeda, H. Yoshida, S. Sekine, Y. Fujita, T. Suzuki, and S. Otaka. A 950-MHz rectifier circuit for sensor network tags with 10-m distance. *IEEE J. of Solid-State Circuits*, 41(1):35–41, 2006.
- [113] Nagaveni Vamsi, V Priya, Ashudeb Dutta, and Shiv Govind Singh. A 1V, -26dBm sensitive auto configurable mixed converter mode rf energy harvesting with wide input range. In *International Symposium on Circuits and Systems (ISCAS)*, pages 1534–1537. IEEE, 2016.
- [114] Roni Vardi, Amir Goldental, Shira Sardi, Anton Sheinin, and Ido Kanter. Simultaneous multi-patch-clamp and extracellular-array recordings: Single neuron reflects network activity. *Scientific reports*, 6:36228, 2016.
- [115] Woradorn Wattanapanitch and Rahul Sarpeshkar. A low-power 32-channel digitally programmable neural recording integrated circuit. *Biomedical Circuits and Systems, IEEE Transactions on*, 5(6):592–602, 2011.
- [116] Michael T Wolf, Jorge G Cham, Edward A Branchaud, Grant H Mulliken, Joel W Burdick, and Richard A Andersen. A robotic neural interface for autonomous positioning of extracellular recording electrodes. *The International Journal of Robotics Research*, 2009.
- [117] Stephen Wolfram. *A new kind of science*, volume 5. Wolfram media Champaign, IL, 2002.
- [118] SS Yarandi and S Srinivasan. Diabetic gastrointestinal motility disorders and the role of enteric nervous system: current status and future directions. *Neurogastroenterology & Motility*, 26(5):611–624, 2014.
- [119] Jun Yi, Wing Hung Ki, and Chi Ying Tsui. Analysis and design strategy of UHF micro-power CMOS rectifiers for micro-sensor and RFID applications. *IEEE Trans. on Circuits and Systems I: Regular Papers*, 54(1):153–166, 2007.
- [120] Tat-Kwan Yu, Sejal Chheda, J Ko, Mark Robertson, Aykut Dengi, and

- Ed Travis. A two-dimensional low pass filter model for die-level topography variation resulting from chemical mechanical polishing of ild films. In *Electron Devices Meeting, 1999. IEDM'99. Technical Digest. International*, pages 909–912. IEEE, 1999.
- [121] Pu-Ming Zhang, Jin-Yong Wu, Yi Zhou, Pei-Ji Liang, and Jing-Qi Yuan. Spike sorting based on automatic template reconstruction with a partial solution to the overlapping problem. *Journal of neuroscience methods*, 135(1):55–65, 2004.
- [122] Xianglilan Zhang, Jiping Sun, and Zhigang Luo. One-against-all weighted dynamic time warping for language-independent and speaker-dependent speech recognition in adverse conditions. *PloS one*, 9(2):e85458, 2014.